

Kapitel 2

Statistische Methoden

2.1 Lage- und Streuungsmaße

2.1.1 Kindersegen

15 Frauen wurden nach der Anzahl ihrer Kinder befragt.

Nummer Probandin	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Anzahl Kinder	0	2	2	1	2	1	0	1	3	4	0	1	2	0	1

Aufgaben:

- Ermitteln Sie den Mittelwert, Median sowie Modalwert.
- Wie hoch ist die Varianz der Stichprobe?
- Wie hoch ist die Standardabweichung der Stichprobe?

2.1.2 Morbus Crohn

In einer kleinen Umfrage wurden 11 Patienten mit Morbus Crohn, einer chronisch-entzündlichen Darmerkrankung gefragt, welchen medikamentösen Wirkstoff sie zurzeit primär einnehmen. Die Antworten lauteten: Azathioprin, Budesonid, Mesalazin, Budesonid, Prednisolon, Mesalazin, Mercaptopurin, Mercaptopurin, Sulfasalazin, Mesalazin, Prednison.

Aufgabe:

Ermitteln Sie den Mittelwert, Median sowie Modalwert.

2.2 Wahrscheinlichkeiten und Kombinatorik

2.2.1 Würfellaune

Vor Ihnen liegt ein handelsüblicher Würfel mit den Zahlen von 1 bis 6.

Aufgaben:

Wie hoch ist die Wahrscheinlichkeit, dass Sie folgende Zahlen bzw. Zahlenkombinationen würfeln?

- a) Eine Sechs
- b) Eine Sechs oder eine Drei.
- c) Zwei Sechsen hintereinander.
- d) Die Zahlenfolge eins bis sechs.
- e) Die Zahlenfolge sechs bis zwei.
- f) Sechs Sechsen hintereinander.
- g) In sechs Würfeln die Zahlen eins bis sechs zu würfeln (unabhängig von der Reihenfolge).

Wie hoch ist die Wahrscheinlichkeit bei einem Wurf mit zwei Würfeln

- a) die Augenzahl 11 zu erzielen,
- b) höchstens die Augenzahl 4 zu würfeln,
- c) höchstens die Augenzahl 11 zu erzielen?

2.2.2 Wellness-Urlaub

Folgende Ausgestaltungen von Lotterien liegen Ihnen vor:

1. Aus einer Lostrommel mit 50 Kugeln (Zahlen 1–50) werden 5 Zahlen nacheinander gezogen (ohne Zurücklegen), wobei die Reihenfolge der Zahlen entscheidend ist.
2. Aus einer Lostrommel mit 40 Kugeln (Zahlen 1–40) werden 5 Zahlen nacheinander gezogen (mit Zurücklegen), wobei die Reihenfolge der Zahlen entscheidend ist.
3. Aus einer Lostrommel mit 50 Kugeln (Zahlen 1–50) werden 7 Zahlen nacheinander gezogen (ohne Zurücklegen). Die Reihenfolge der Ziehung ist nicht entscheidend.
4. Aus einer Lostrommel mit 60 Kugeln (Zahlen 1–60) werden 6 Zahlen nacheinander gezogen (mit Zurücklegen). Die Reihenfolge der Ziehung ist nicht entscheidend.

Bei richtiger Kombination winkt dem Gewinner bzw. der Gewinnerin ein Wellness-Urlaub.

Aufgabe:

Ermitteln Sie die jeweilige Wahrscheinlichkeit, die Wellness-Reise bei Teilnahme an der Lotterie zu gewinnen. Stellen Sie dabei zunächst die Formel für die Kalkulation der Wahrscheinlichkeit auf. Falls die Teilnahmegebühr bei allen vier Lotterien gleich ist, bei welcher würden Sie am ehesten mitspielen?

2.2.3 Epidemiologische Maßzahlen und Wahrscheinlichkeit

Aufgrund seiner familiären Erkrankungsgeschichte nimmt ein fünfzigjähriger Mann, der bislang keine Symptome aufweist, an einer Darmkrebs-Vorsorgeuntersuchung teil. Die Prävalenz in seiner Risikogruppe sei 10 %. Die Sensitivität der Untersuchung betrage 96 % und die Spezifität 97 %.

Aufgaben:

- Wie hoch ist die Wahrscheinlichkeit bei einem Positivbefund, dass der untersuchte Mann Darmkrebs hat ($p(K+|T+)=PPV$)? Wie hoch ist die Wahrscheinlichkeit, dass er keinen Darmkrebs hat?
- Mit welcher Wahrscheinlichkeit hat der Patient keinen Darmkrebs bei einem Negativbefund ($p(K-|T-)$), mit welcher Wahrscheinlichkeit hat er trotz eines Negativbefundes einen Darmkrebs?

2.3 Binomial- und Hypergeometrische Verteilung**2.3.1 Einführung: Medizintechnikgroßhändler**

Ein Medizintechnikgroßhändler bekommt eine Lieferung von drei Herzschrittmachern. Der Großhändler möchte die medizinischen Geräte an Krankenhäuser in der Region weiterverkaufen, vorher muss die medizintechnische Abteilung jedoch die Funktionstüchtigkeit der Herzschrittmacher überprüfen. Es seien H_1 , H_2 , H_3 die drei Ereignisse, dass der erste, zweite bzw. dritte Herzschrittmacher defekt ist. Beschreiben Sie mit Hilfe dieser Variablen und geeigneter Mengenoperationen die nachstehenden Ereignisse. Veranschaulichen Sie die Ereignisse jeweils durch ein Mengendiagramm.

Aufgaben:

- Alle Herzschrittmacher sind funktionstüchtig.
- Alle Herzschrittmacher sind defekt.
- Mindestens ein Herzschrittmacher ist defekt.
- Genau zwei Herzschrittmacher sind defekt.
- Höchstens ein Herzschrittmacher ist defekt.

2.3.2 *Krebs in einer Population*

In einer Population sind 5 % der Bevölkerung an Krebs erkrankt.

Aufgaben:

Wie groß ist die Wahrscheinlichkeit, dass ... (jeweils mit Zurücklegen):

- a) 5 zufällig ausgewählte Personen an Krebs erkrankt sind?
- b) von 10 zufällig ausgewählten Personen keiner an Krebs erkrankt ist?
- c) von 5 zufällig ausgewählten Personen höchstens 2 an Krebs erkrankt sind?
- d) höchstens 2 von 3 zufällig ausgewählten Personen an Krebs erkrankt sind?
- e) genau einer von zufällig ausgewählten 6 Personen an Krebs erkrankt sind?

2.3.3 *Fehlerhafte Ware*

Ein Krankenhaus erhält wöchentlich eine 68 Kisten umfassende Lieferung an medizinischem Zubehör. Der Lieferant hat bei der Werkskontrolle dieser Lieferung eine Fehlertoleranz von 4 Kisten. Das Krankenhaus kontrolliert aus jeder Lieferung 10 zufällig ausgewählte Kisten auf Vollständigkeit und korrekte Bestückung. Mit dem Lieferanten ist vereinbart, dass falls bei der Kontrolle mehr als eine fehlerhafte Kiste gefunden wird, die Lieferung zurückgegeben wird.

Aufgaben:

- a) Wie hoch ist bei dem vom Krankenhaus gewähltem Kontrollverfahren die Wahrscheinlichkeit, dass eine an sich den Fehlertoleranzen entsprechende Lieferung zurückgewiesen wird, wenn jede Kiste nach der Überprüfung in den Stapel zurückgelegt wird? (Ziehen mit Zurücklegen)
- b) Wie hoch ist die Wahrscheinlichkeit, wenn die einmal überprüften Kisten gesondert aufbewahrt werden? (Ziehen ohne Zurücklegen)

2.3.4 *Qualitätskontrolle*

Ein kleines Pharmaunternehmen stellt täglich 300 Packungen mit jeweils 5 Blistern à 10 Tabletten eines bestimmten Medikamentes her. Aus technischen Gründen ist es möglich, dass eine Packung mit Arzneimitteln am Fließband unzureichend bestückt wird. Beispielsweise könnten ein Blister in der Packung oder aber einzelne Tabletten in einem Blister fehlen. Aus diesen Gründen werden die Arzneimittelpackungen einer Qualitätskontrolle unterzogen, bevor sie abends an den Großhändler versandt werden. Dazu werden jeden Abend zufällig ausgewählte Packungen auf ihren Inhalt hin untersucht. Im Rahmen der Qualitätskontrolle werden die bereits geprüften Päckchen gesondert gelagert.

Aufgaben:

- a) Wie hoch ist die Wahrscheinlichkeit höchstens zwei fehlerhafte Päckchen zu finden, wenn 10 % der Produktion überprüft werden und angenommen wird, dass insgesamt 10 Päckchen nicht einwandfrei bestückt sind?
- b) Wie hoch ist die Wahrscheinlichkeit kein fehlerhaftes Päckchen zu finden, wenn 5 % der Produktion überprüft werden und angenommen wird, dass insgesamt 1 % Päckchen nicht einwandfrei bestückt sind?
- c) Nehmen Sie an, dass die Produktion nicht an den Großhändler geschickt wird, wenn bei einer Stichprobe von 40 Päckchen mehr als ein Päckchen von der Maschine falsch bestückt worden ist. Der Hersteller, der die Produktionsanlage geliefert hat, gibt in seiner Werbebroschüre an, dass bei „nur“ 1 von 75 Päckchen eine fehlerhafte Zusammensetzung vorkommen kann. Wie hoch ist die Wahrscheinlichkeit, dass die Lieferung nicht an den Großhändler geschickt wird?

2.4 Normalverteilung

2.4.1 *Blutzucker*

Bei einer Gruppe von 500 zufällig ausgewählten Frauen wurde der Blutzuckerspiegel gemessen. Der Durchschnitt in der Gruppe betrug 110 mg/dl mit einer Standardabweichung von 30. Es wird eine Normalverteilung der Stichprobe angenommen.

Nutzen Sie für die folgenden Aufgaben die z-Tabelle (siehe Anhang A).

Aufgaben:

- a) Wie hoch ist die Wahrscheinlichkeit einen Blutzuckerspiegel von weniger als 145 mg/dl zu haben?
- b) Wie hoch ist die Wahrscheinlichkeit, einen Blutzuckerspiegel von mehr als 180 mg/dl zu haben?
- c) Wie hoch ist die Wahrscheinlichkeit einen Blutzuckerspiegel von genau 130 mg/dl zu haben?
- d) Betrachten wir die 25 Frauen mit dem niedrigsten Blutzuckerspiegel: welchen Blutzuckerspiegel werden diese wahrscheinlich nicht überschreiten?
- e) Betrachten Sie die 10 Frauen mit dem höchsten Blutzuckerspiegel: welchen Blutzuckerspiegel werden diese wahrscheinlich nicht unterschreiten?
- f) Zwischen welchen Werten werden 80 % der Frauen wahrscheinlich liegen?
- g) Welche Intervalle können in einer normalverteilten Stichprobe ohne Blick auf die z-Tabelle schnell, bei Kenntnis des Mittelwertes sowie der Standardabweichung, approximativ ermittelt werden? Veranschaulichen Sie Ihr Ergebnis anhand einer graphischen Darstellung der Dichtefunktion der Verteilung. Welche Werte können Sie demnach direkt für die vorliegende Stichprobe angeben?

2.4.2 Hämoglobinwerte

Vor einer Blutspende wird neben dem Blutdruck und der Körpertemperatur auch der Hämoglobinwert (Hb-Wert), auch vereinfacht Eisenwert genannt, gemessen. Normalerweise liegt der Hb-Wert bei Frauen zwischen 12 und 16 g/dl, bei Männern zwischen 14 und 18 g/dl. Um Blutspenden zu können, müssen Frauen mindestens einen Wert von 12,5 g/dl aufweisen und Männer einen Wert von 13,5 g/dl. Mit einem Eisenwert von über 18 g/dl werden die potentiellen Spender und Spenderinnen i. d. R. nicht zur Blutspende zugelassen, sondern erst einmal weitere Untersuchungen veranlasst. (Herzlichen Dank an den Blutspendedienst der Medizinischen Hochschule Hannover (MHH) für diese Auskunft).

In einer Kohorte von 450 Männern und 500 Frauen, die gerne Blutspenden möchten, liegt der durchschnittliche Eisenwert bei 16 bzw. 14 g/dl mit einer Standardabweichung von jeweils 1,5. Es wird eine Normalverteilung angenommen.

Nutzen Sie für die folgenden Aufgaben die z-Tabelle (siehe Anhang A).

Aufgaben:

- Wie hoch ist die Wahrscheinlichkeit, als Mann einen Hb-Wert von weniger als 14 g/dl zu haben?
- Wie hoch ist die Wahrscheinlichkeit, als Frau einen Hb-Wert von über 15 g/dl zu haben?
- Wie hoch ist die Wahrscheinlichkeit, als Frau einen Hb-Wert von über 12,5 g/dl zu haben?
- Wie hoch ist die Wahrscheinlichkeit als Mann einen Hb-Wert von genau 16 g/dl bzw. als Frau einen Hb-Wert von genau 14 g/dl zu haben?
- Betrachten wir die 50 Frauen mit dem niedrigsten Eisenspiegel: welchen Hb-Wert werden diese wahrscheinlich nicht überschreiten?
- Betrachten wir die 20 Männer mit dem höchsten Eisenspiegel: welchen Hb-Wert werden diese wahrscheinlich nicht unterschreiten?
- Zwischen welchen Werten werden 60 % der potentiellen Blutspender bzw. Blutspenderinnen wahrscheinlich liegen?
- Werden mehr Frauen oder mehr Männer zur Spende zu gelassen? Berechnen Sie die wahrscheinliche Anzahl.

2.4.3 Körpergröße

In einer Gruppe von 20 zufällig ausgewählten männlichen Studierenden messen Sie die Körpergröße. Im Ergebnis sind die Studenten durchschnittlich 185 cm groß mit einer Standardabweichung von 10 cm.

Aufgabe:

- Welchen Wert werden 90 % der Studenten wahrscheinlich nicht überschreiten? Unterstellen Sie eine t-Verteilung der Werte mit $n-1$ Freiheitsgraden. (Nutzen Sie hierzu die t-Tabelle im Anhang A.)

- b) Nehmen Sie an, dass die Gruppe nicht aus 20, sondern aus 500 Studierenden besteht. Mittelwert und Standardabweichung bleiben unverändert. Ändert sich das Ergebnis aus Aufgabenteil a)?

2.5 Poissonverteilung

2.5.1 Unfälle in einer Versicherungskohorte

Stellen Sie sich vor, Sie sind Versicherungsmathematiker in einer privaten Krankenversicherung mit insgesamt 100.000 Versicherten, deren Verteilung der deutschen Bevölkerung entspricht. Ihre Vorgesetzte bittet Sie, bei einem selten vorkommenden Unfall A anzugeben, wie wahrscheinlich es ist, dass sich ein oder mehrere Fälle in ihrer Versichertenkohorte ereignen. Ihnen ist bekannt, dass Unfall A in einem Jahr durchschnittlich in nur 4 von 100.000 Fällen auftritt.

Nutzen Sie für die folgenden Aufgaben die Tabelle zur Poisson-Verteilung (siehe Anhang A).

Aufgaben:

Wie wahrscheinlich ist es, dass aus ihrer Versicherungskohorte

- a) niemand diesen Unfall erleidet,
- b) weniger als 5 Personen diesen Unfall erleiden,
- c) mehr als 6 diesen Unfall erleiden,
- d) 7 oder 8 Personen diesen Unfall erleiden,
- e) 10 oder mehr Personen diesen Unfall erleiden?

2.5.2 Seltene Erkrankung

Sie sind immer noch Versicherungsmathematiker ... jedoch sind Sie jetzt in eine größere Krankenversicherung mit insgesamt 200.000 Versicherten gewechselt. Deren Verteilung entspricht weiterhin der deutschen Bevölkerung. Ihr Vorgesetzter bittet Sie zu ermitteln, wie hoch die Wahrscheinlichkeit ist, dass eine seltene Erkrankung S, die nicht genetisch bedingt ist und in jeder Altersklasse auftreten kann, in Ihrer Versichertenkohorte auftritt. Ihnen ist bekannt, dass S in einem Jahr durchschnittlich in nur 1 von 1.000.000 Fällen auftritt.

Nutzen Sie für die folgenden Aufgaben die Tabelle zur Poisson-Verteilung (siehe Anhang A).

Aufgaben:

Wie wahrscheinlich ist es, dass aus ihrer Versicherungskohorte

- a) niemand diese Krankheit bekommt,
- b) weniger als 2 Personen diese Krankheit bekommen,

- c) mehr als 3 Personen diese Krankheit bekommen,
- d) 2 oder 3 Personen diese Krankheit bekommen,
- e) 4 oder mehr Personen diese Krankheit bekommen?

2.6 Statistische Tests

2.6.1 Einseitig versus zweiseitige Tests

Aufgabe:

Erläutern Sie kurz den grundsätzlichen Unterschied zwischen einem einseitigen und einem zweiseitigen Test und verdeutlichen Sie den Unterschied anhand einer Graphik.

2.6.2 Diabetes mellitus Typ II

In der folgenden Tabelle sind die Kosten von 36 Diabetes mellitus Typ II-Patienten aufgelistet, die im Laufe eines Jahres angefallen sind. Alle Patienten sind eingeschrieben in ein Disease Management Programm (DMP) und wurden zufällig im Rahmen der Programme ausgewählt.

3.000	3.000	3.000	8.000	5.000	2.000
2.000	15.000	7.000	500	12.000	3.000
1.000	3.500	5.000	4.000	1.500	2.500
10.000	4.000	2.500	8.000	1.000	3.000
1.500	5.000	1.000	2.500	2.000	3.500
2.000	500	1.000	4.000	2.500	4.000

Aufgaben:

- a) Geben Sie den Minimal- und Maximalwert sowie Mittelwert, Median und Modalwert der Stichprobe an.
- b) Wie hoch ist die Varianz der Stichprobe? Wie hoch ist die Standardabweichung?
- c) Ihnen ist aus einer anderen Quelle bekannt, dass Diabetiker, die nicht in ein DMP eingeschrieben sind, jährlich im Erwartungswert Kosten in Höhe von 4.500 € verursachen (μ_{nDMP}). Sie möchten nun anhand Ihrer Stichprobe überprüfen, ob sich die Kosten von Diabetikern, die in ein DMP eingeschrieben sind, signifikant ($p < 0,05$) in der Grundgesamtheit von denen unterscheiden, die nicht in ein DMP eingeschrieben sind. Stellen Sie zunächst Ihre Hypothesen auf. Bestimmen Sie

dann das Konfidenzintervall und verdeutlichen Sie Ihr Ergebnis anhand einer Graphik. Gehen Sie von einer Normalverteilung der Zufallsvariable aus. (Nutzen Sie hierzu die t-Tabelle im Anhang A.)

- d) Inwiefern ändern sich Ihre Ergebnisse aus Aufgabe c), wenn Sie nicht eine Stichprobe von 36 sondern stattdessen 151 Patienten betrachten? (Annahme: Lage- und Streuungsmaße bleiben gleich.) Berechnen Sie das Konfidenzintervall und verdeutlichen Sie Ihr Ergebnis anhand einer Graphik. Gehen Sie wiederum von einer Normalverteilung der Zufallsvariable aus. (Nutzen Sie hierzu die t-Tabelle im Anhang A.)
- e) Diskutieren Sie kurz die Annahme der Normalverteilung.

2.6.3 Sind WiWi-Studierende wehleidiger?

51 zufällig ausgewählte Studierende der Wirtschaftswissenschaften (WiWi) wurden gefragt, wie häufig sie im letzten Jahr zum Arzt gegangen sind. Im Ergebnis haben die WiWi-Studierenden durchschnittlich 8-mal innerhalb des letzten Jahres einen Arzt aufgesucht. Im Vergleich gingen laut Statistik Studierende anderer Fachrichtungen im Durchschnitt 6-mal im Jahr zum Arzt. Die Standardabweichung beträgt 5,9. Für die untersuchte Gruppe der Studierenden wird Normalverteilung unterstellt.

Aufgaben:

- a) Testen Sie zum Konfidenzniveau von 90 %, ob sich die Anzahl der Arztbesuche der WiWi-Studierenden signifikant von der Anzahl der Arztbesuche der Grundgesamtheit der Studierenden anderer Fachrichtungen unterscheidet. Formulieren Sie dabei zunächst Ihre Hypothesen.
- b) Inwiefern ändert sich das Ergebnis aus Aufgabeteil a), wenn Sie zu einem Konfidenzniveau von 99 % testen?

2.6.4 Fehlzeiten

Eine Befragung von 20 Personen hinsichtlich der Fehlzeiten am Arbeitsplatz pro Jahr ergab folgendes Ergebnis:

Geschlecht	F	F	F	M	F	M	F	F	M	M	F	F	M	F	F	M	F	F	M	F
Anzahl Fehlstunden	49	38	27	33	57	44	64	58	24	51	47	39	37	62	21	34	26	46	54	36

Aufgabe:

Testen Sie mit dem Mann-Whitney-U-Test, ob sich die Fehlzeiten zwischen Frauen (F) und Männern (M) bei einer Fehlerwahrscheinlichkeit von 5 % signifikant voneinander unterscheiden. (Hinweis: U_{krit} bei 0,05 mit $m=7$ und $n=13$ entspricht einem Wert von 20).

2.6.5 Korrekte Verteilung?

Hinsichtlich des Auftretens einer neuartigen Krankheit wurden folgende altersabhängige Beobachtungen gemacht:

Altersklasse i	Registrierte Erkrankte
< 10 Jahre	15
10-30 Jahre	40
30-50 Jahre	30
> 50 Jahre	15

Ein Mediziner behauptet nun, die Krankheit erforscht zu haben und in der Lage zu sein, folgende Prävalenzen für die einzelnen Altersklassen angeben zu können:

Altersklasse i	Prävalenz
< 10 Jahre	20%
10-30 Jahre	30%
30-50 Jahre	30%
> 50 Jahre	20%

Aufgabe:

Testen Sie mit dem Chi-Quadrat-Anpassungstest bei einem Signifikanzniveau von 0,05, ob das Verteilungsmodell des Mediziners den Krankheitsverlauf richtig beschreibt. (Hinweis: $\chi^2_{\text{df}; 1-\alpha} = \chi^2_{3; 0,95} = 7,815$).

2.7 Korrelation und Regression**2.7.1 Korrelation versus Regression****Aufgabe:**

Erläutern Sie kurz den Unterschied zwischen einer Korrelationsanalyse und einer einfachen linearen Regressionsanalyse.

2.7.2 Rauchen & Alkohol

Im Rahmen einer Studie wurden bei 13 Probanden die täglich gerauchte Anzahl an Zigaretten (x) und die tägliche getrunkene Menge an Alkohol in Zentilitern (y) erhoben.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	3	1	5	3	7	2	7	8	6	6	2	5	4
y_i	30	50	60	10	50	30	20	10	40	70	80	60	90

Aufgabe:

Prüfen Sie, ob es einen Zusammenhang zwischen diesen beiden Merkmalen gibt, indem Sie den Korrelationskoeffizienten berechnen.

2.7.3 Die Störgröße im Regressionsmodell

Aufgabe:

Wozu dient die Störgröße im Regressionsmodell?

2.7.4 Zappelphillip

Eine neue Studie soll untersuchen, ob es einen Zusammenhang zwischen der Anzahl der eingenommenen Retalin-Tabletten pro Tag und der Aufmerksamkeitsspanne eines Kindes gibt. Nachfolgend sind die Daten von 10 untersuchten Kinder angegeben:

Kind i	1	2	3	4	5
Retalin-Tabletten pro Tag (x)	2	0,3	3,1	2,8	2,6
Aufmerksamkeitsspanne in Minuten (y)	40	50	90	75	80
Kind i	6	7	8	9	10
Retalin-Tabletten pro Tag (x)	0,5	1,2	0,5	0,9	1,7
Aufmerksamkeitsspanne in Minuten (y)	20	45	30	42	61

Aufgaben:

- a) Ermitteln Sie die Regressionsgerade und zeichnen Sie sowohl die Daten (als Punkte) als auch die Regressionsgerade in ein geeignetes Koordinatensystem ein.
- b) Berechnen Sie die Residuen für jedes Kind.
- c) Berechnen Sie das Bestimmtheitsmaß R^2 . Welche Aussage über die Güte der Regression können Sie mittels des Bestimmtheitsmaßes treffen?

Übungen zu Public Health und Gesundheitsökonomie
Prenzler, A.; Graf von der Schulenburg, J.-M.; Zeidler, J.
2010, XIV, 117 S., Softcover
ISBN: 978-3-642-13504-0