ABCs zu Data Analytics II

(für Tablet mit Lösungen)

Prof. Dr. Wolfgang Bischof

TH Augsburg

27. Januar 2026

ABC-Spiel (1)

Aufgabe 1.1

Mit welchem Test kann am besten signifikant nachgewiesen werden, dass ihre Daten normalverteilt sind?

A: Shapiro-Wilk

B: Normalscores-Test

C: Lillifors-Test

D: mit keinem

Lösung: D, da prinzipiell mit keinem Test signifikant nachgewiesen werden kann, dass Daten normalverteilt sind. Es kann nur signifikant nachgewiesen werden, dass Daten nicht normalverteilt sind.

Aufgabe 1.2

Auf dem Plärrer steht bei der Losbude des roten Kreuzes: „Jedes fünfte Los gewinnt.“. Beim privaten Anbieter Teddy-Bär steht: „Die Chance auf einen Gewinn beträgt eins zu fünf.“. Wie ist das Chancenverhältnis (Odds Ratio) zwischen Rot-Kreuz- und Teddybären-Bude.

A: $\frac{4}{5}$ B: $\frac{5}{6}$ C: 1 D: $\frac{5}{4}$ E: $\frac{6}{5}$

Lösung: D, da Chancenverhältnis = (Chance RK) : (Chance TB) = $\frac{1/5}{4/5} = \frac{5}{4}$

Aufgabe 1.3 Eine englische Lady behauptet, dass sie erkennen könne, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde. Ein deutsche Versuchsperson hat dagegen keine Ahnung vom Tee-trinken und rät, ob Tee oder Milch zuerst in die Tasse gegossen wurde. Es werden nun jeweils zehn Tassen vorbereitet, bei denen die Reihenfolge von Tee und Milch zufällig ist. Mit welchem Test kann die Lady nachweisen, dass sie die Reihenfolge signifikant besser erkennen kann als die deutsche Versuchsperson.

A: Exakter Test von Fisher

B: Einseitiger Binomialtest

C: Zweiseitiger Binomialtest

D: Zweistichproben t-Test

E: Zweistichproben Tee-Test

Lösung: A, da es sich um das Testen von zwei Erfolgswahrscheinlichkeiten in zwei zwei Stichproben handelt.

Aufgabe 1.4

Wahr (A), falsch (B) oder „Da müsste ich raten.“ (C)?

(i) Mit dem exakten Test von Fisher lässt sich signifikant nachweisen, dass zwei Merkmale voneinander unabhängig sind.

Lösung: B, da mit dem exakten Test von Fisher nur die Unabhängigkeit signifikant widerlegt, nicht aber bewiesen werden kann.

(ii) Muriel Bristol behauptet, dass sie erkennen kann, ob zuerst Milch oder Tee in eine Tasse gegossen wurde. Nun werden ihr acht Tassen präsentiert, von denen in zufälliger Reihenfolge bei vier zuerst Milch und dann Tee und bei den anderen vier zuerst Tee und dann Milch in die Tasse gegossen wurde. Sie darf alle Tassen testen und muss am Ende des Tests sagen, bei welchen vier zuerst Milch eingegossen wurde. Nehmen Sie nun an, dass sie alle vier Tassen richtig identifiziert hat.

Der p-Wert des Testproblems:

beträgt $Hyp_{4,4,4}(4) =$`dhyper(4,4,4,4)`$\approx 1.4\%$.

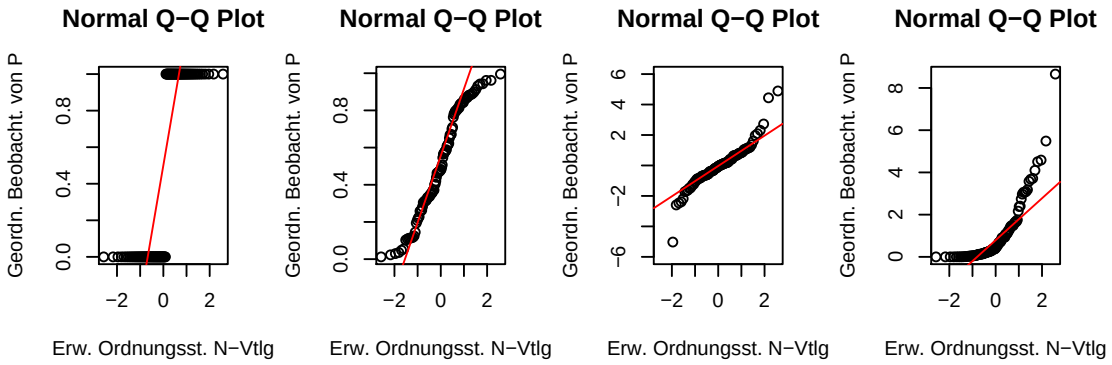
Bemerkung: Das Problem ist als „Lady tasting tea“-Problem bekannt.

Lösung: A, da die Wsk. unter der Nullhypothese („Raten“) vier (oder mehr, was nicht geht) richtige Tassen zu bestimmen, $Hyp_{4,4,4}(4) = \frac{1}{\binom{8}{4}} = \frac{1}{70}$ beträgt.

Hinweis: Das Problem kann auch mit dem exakten Test von Fisher gelöst werden, in dem man die beiden Merkmale „Lady schätzt Milch vor Tee“ und „Milch wurde vor Tee eingegossen“ auf einseitige Unabhängigkeit prüft, etwa mit R: `tea <- matrix(c(4,0,0,4), nrow=2); fisher.test(tea, alternative = "greater")`.

ABC-Spiel (2)

Aufgabe 2.1 Seien $X_1, \ldots, X_{100}$ u.i. gemäß P verteilt. In den folgenden vier Q-Q-Plots werden jeweils die beobachteten Ordnungsstatistiken $X_{\uparrow 1}(\omega), \ldots, X_{\uparrow 100}(\omega)$ gegen die erwarteten Ordnungsstatistiken von 100 standardnormalverteilten Zufallsvariablen aufgetragen. Entscheiden Sie zu jedem der vier Plots, von welcher Verteilung P die beobachteten Ordnungsstatistiken stammen?



A: $N_{0,1}$ C: χ^2_1 E: t_2

B: $U_{[0,1]}$ D: $B_{1;0,5}$ F: $U_{\{1,2,3,4,5,6\}}$

Lösung:

(i) D, da die $B_{1,0.5}$-Verteilung die einzige der angegebenen Verteilungen ist, die nur die Werte 0 und 1 annimmt.

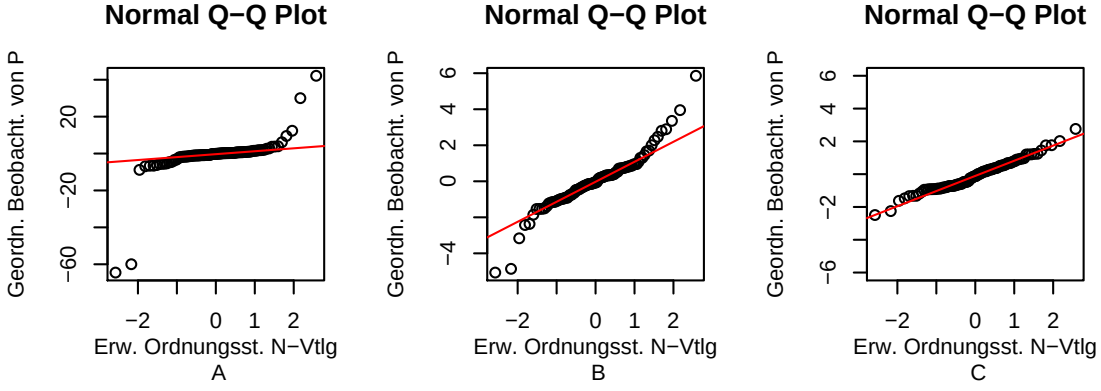
(ii) B, da die $U_{[0,1]}$-Verteilung die einzige der angegebenen Verteilungen ist, die nur Werte zwischen 0 und 1 annimmt und auch Zwischenwerte annehmen kann.

(iii) E, da aufgrund der auftretenden Werte nur die $N_{0,1}$- und t_2-Verteilung in Frage kommen und die $N_{0,1}$-Verteilung ausscheidet, da der Graph aufgrund der Abweichungen an den Rändern der Gerade darauf schließen lässt, dass die Verteilungsränder der Beobachtungswerte viel stärker besetzt sind als bei der Normalverteilung, genau wie es bei einer t_2-Verteilung der Fall ist.

(iv) C, da die χ^2_1-Verteilung die einzige der angegebenen Verteilungen ist, die Werte zwischen 0 und größer als 10 annimmt.

Aufgabe 2.2

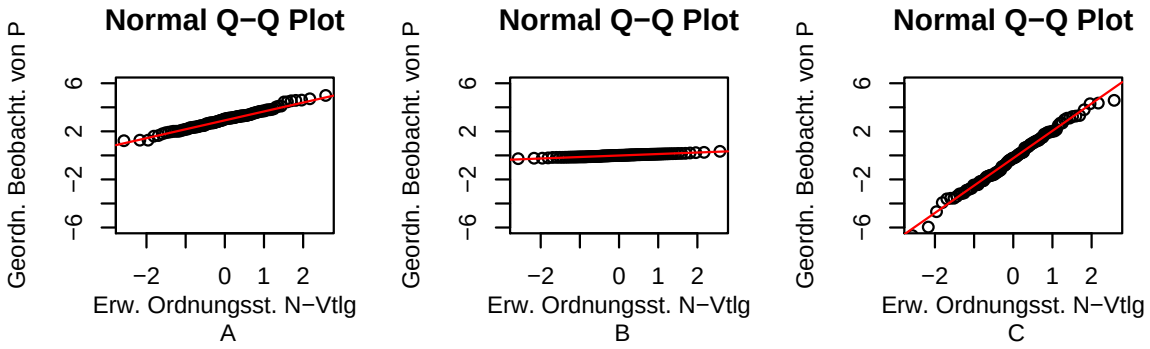
Sei die Situation aus Aufg. 2.1 gegeben. Die drei unteren Plots stammen von den Verteilungen t_1, t_3 und t_{100}. Welcher Plot gehört zu t_1?



Lösung: A, da die Streuung in dieser Grafik bei weitem die größte ist (Werte liegen zwischen -60 und 40).

Aufgabe 2.3

Sei die Situation aus Aufg. 2.1 gegeben. Die drei unteren Plots stammen von den Verteilungen $N_{3,1}$, $N_{0,0.01}$ und $N_{0,4}$. Welcher Plot gehört zu $N_{0,4}$?



Lösung: C, da die Beobachtungswerte symmetrisch zur 0 liegen und im 3σ-Bereich, d.h. von -6 bis 6.

AB-Wert:

Erw.wert	Max	95% bei einem Affe	95% bei 30 Affen
1.3	6	3	1.6

ABC-Spiel (3)

Aufgabe 3.1 Bei der letzten US-Präsidentenwahl hatte Kandidat A 45% erzielt, Amtsinhaber O 55%. Nun treten A und O wieder an. Kurz vor der neuen Wahl soll durch eine Wahlumfrage ermittelt werden, ob sich die Wähler nun anders als bei der letzten Wahl entscheiden. Welcher Test ist dazu am besten geeignet?

A: Normalscores-Test
B: t-Test
C: Wilcoxon-Rangsummentest
D: Binomialtest
E: Exakter Test von Fisher
F: Wilcoxon-Vorzeichen-Rang-Test

Lösung: D, und zwar der zweiseitige Binomialtest $p_O = 0.55$ gg. $p_O \neq 0.55$. Der χ^2-Anpassungstest wäre zwar auch anwendbar, ist aber nur ein approximatives Verfahren und daher dem Binomialtest unterlegen.

Aufgabe 3.2

In einem Dorf in den Stauden westlich von Augsburg gibt es ca. 20% Rennradler und 80% normale Menschen. Nun beginnt eine zufällig ausgewählte Gruppe des Dorfes jeden Morgen zum Frühstück „Bischofs flottes Müsli.“ zu genießen. Nach einem Monat will der Bürgermeister wissen, ob das Müsli wirklich so gut ist, wie der Werbespruch verheißt. Dazu wählt er zufällig 10 Personen aus der Gruppe der Müsliesser und 10 Personen aus der Gruppe, die normal gefrühstückt haben, aus und schaut, wie weit jede Person in einer halben Stunde mit dem Rad fahren kann.

Welchen Test wählt er?

A: Normalscores-Test
B: Approximativen Zweistichproben-t-Test
C: Wilcoxon-Rangsummentest
D: t-Test für gepaarte Beobachtungen
E: Exakter Test von Fisher
F: Wilcoxon-Vorzeichen-Rang-Test

Lösung: C, da die Daten sicher nicht normalverteilt sind (zweigifplige Verteilung der Radstrecke).

Aufgabe 3.3

Vier Chinesen treten gegen vier Bayern im Schafkopf an. Nach einem langen Abend, an dem die Teams immer wieder gewechselt werden, wurden folgende Punktesummen erzielt:

Name	Ping	Franz	Rosi	Su	Sepp	Hu	Vroni	Li
Punkte	56	32	47	31	50	0	35	17

Welcher Wert ergibt sich für die W-Statistik der Chinesen, wenn - wie bei der W-Statistik üblich - die Platzziffer 1 an denjenigen mit den wenigsten Punkten vergeben wird?

A: 22 B: 14 C: 16 D: 20

Lösung: B, da 1+2+3+8=14.

Aufgabe 3.4

Welcher Wert ergibt sich für die U-Statistik der Chinesen?

A: 12 B: 8 C: 4 D: 22

Lösung: C, da Ping der einzige ist, der mehr Punkte als irgendwelche Bayern erzielt hat. Es haben vier Bayern weniger Punkte als er erzielt.
Alternativ: $U = W - \frac{5 \cdot 4}{2} = 14 - 10 = 4$.

Aufgabe 3.5

Betrachten Sie die Situation aus den Aufgaben 3.3 und 3.4. Die U- und W-Statistik unterscheiden sich unabhängig von den Punktesummen nur durch eine Konstante.

A: wahr B: falsch C: Da müsste ich raten.

Lösung: A, da laut Vorlesung gilt: $W - U = \frac{k(k+1)}{2}$, wobei k die Anzahl der Beobachtungen in der ersten Gruppe sind.

ABC-Spiel (4)

Aufgabe 4.1

In einer Gemeinde werden die monatlichen Bruttoeinkünfte sowie die Religionszugehörigkeit erfasst. Um nachzuweisen, dass Katholiken typischer mehr verdienen als Protestanten ziehen Sie eine Stichprobe aus 100 Katholiken und 80 Protestanten.

Welche Methodik verwenden Sie?

A: Ich verwende den approximativen Zwei-Stichproben-t-Test, weil mehr als 30 Beobachtungen vorliegen.

B: Ich prüfe mit dem Shapiro-Wilk-Test, ob die Daten in jeder Gruppe mit einer Normalverteilungsannahme verträglich sind. Wenn ja, verwenden ich den approximativen Zwei-Stichproben-t-Test. Wenn nicht, logarithmiere ich die Daten und prüfe ob die logarithmierten Daten in jeder Gruppe mit einer Normalverteilungsannahme verträglich sind. Wenn ja, verwende ich den approximativen Zwei-Stichproben-t-Test für die logarithmierten Daten. Wenn nicht, verwende ich den Wilcoxon-Rangsummentest.

C: Ich teste da gar nix, weil ich nicht religiös bin.

D: Ich prüfe mit dem Shapiro-Wilk-Test, ob die Daten in jeder Gruppe mit einer Normalverteilungsannahme verträglich sind. Wenn ja, verwenden ich den approximativen Zwei-Stichproben-t-Test. Wenn nicht, verwende ich den Wilcoxon-Rangsummentest.

Lösung: B, da die t-Tests eine etwas größere Teststärke als die nicht-parametrischen Tests besitzen und deswegen verwendet werden sollten, wenn die Normalverteilungsannahme für die Daten oder für die logarithmierten Daten gerechtfertigt ist.

Aufgabe 4.2

In einer Gemeinde werden die monatlichen Bruttoeinkünfte sowie die Religionszugehörigkeit erfasst. Um nachzuweisen, dass das Einkommen von der Religionszugehörigkeit abhängt, ziehen Sie eine Stichprobe aus 100 Katholiken, 80 Protestanten, 50 Moslems und 35 Buddhisten. Sie wissen, dass weder die monatlichen Bruttoeinkünfte noch die logarithmierten Bruttoeinkünfte der Moslems normalverteilt sind.

Welchen Test verwenden Sie?

A: Kruskall-Wallis-Test

B: Lilliefors-Test

C: Mehrere approximative Zwei-Stichproben-t-Tests

D: Mehrere Wilcoxon-Zweistichproben-Rangsummen-Tests

E: Wilcoxon-Vorzeichen-Rang-Test

F: Exakter Test von Fisher

Lösung: A, da mehr als zwei Gruppe und ein metrische, nicht normalverteilte Zielgröße vorliegen.

Aufgabe 4.3 Bei der Landtagswahl 2008 in Bayern erzielte die CSU 43.4%, die SPD 18.6%, die Freien Wähler 10.2% die Grünen 9.4% und die FDP 8.0%. Bei einer Umfrage unter 1008 Wahlberechtigten im Mai 2013 ergaben sich folgende Werte: CSU 47%, SPD 20%, Grüne 13%, Freie Wähler 8% und FDP 3%. Mit welchem Test prüfen Sie, ob dieses Ergebnis signifikant von den Wahlergebnissen 2008 abweicht?

A: Appr. Zweistichproben-t-Test

B: χ^2-Anpassungstest

C: Mann-Whitney-U-Test

D: Mehrere Binomialtests

Lösung: B, da die Anteile von mehr als zwei Parteien getestet werden sollen. Mehrere Binomialtests werden der Aufgabe nicht gerecht, da Abhängigkeiten zwischen den Anteilen bestehen und da das Niveau mehrerer Tests nicht dem Gesamtniveau entspricht.

Aufgabe 4.4

Eine Münze wird 10-mal geworfen. Dabei erhält man achtmal Kopf. Es soll nachgewiesen werden, dass die Münze unfair ist. Welchen Test verwenden Sie?

A: χ^2-Anpassungstest

B: Zweiseitiger Binomialtest

C: Approximativer t-Test

D: Wilcoxon-Rangsummentest

Lösung: B, da das Testproblem lautet: $p = 0.5$ gg. $p \neq 0.5$

Aufgabe 4.5

Mit dem χ^2-Anpassungstest kann nachgewiesen werden, dass bei einem potentiell unfairen Würfel die Zahlen 5 und 6 jeweils eine höhere Wahrscheinlichkeit als $\frac{1}{6}$ besitzen.

A: wahr

B: falsch

C: Da müsste ich raten.

Lösung: B, da mit dem χ^2-Anpassungstest nur eine Abweichung von einer Multinomialverteilung nachgewiesen werden kann, nicht aber die Richtung der Abweichung.

ABC-Spiel (5)

Aufgabe 5.1

Es liegen Ihnen 100 Messwerte vor und Sie wollen mit einem χ^2-Anpassungstest prüfen, ob die Messwerte aus einer standard-normal-verteilten Grundgesamtheit stammen. Vor dem Test ordnen Sie die Messwerte vorgegebenen Intervallen zu. Welche der folgenden Einteilungen ist die beste, wenn Sie eine χ^2-Verteilung als Approximation an die Verteilung der Prüfgröße verwenden wollen?

A: $(-\infty, -0.5]$; $(-0.5, 0.5]$; $(0.5, \infty)$

B: $(-\infty, -1]$; $(-1, 1]$; $(1, \infty)$

C: $(-\infty, -0.8]$; $(-0.8, -0.2]$; $(-0.2, 0.2]$; $(0.2, 0.8]$; $(0.8, \infty)$

D: $(-\infty, -3]$; $(-3, -1]$; $(-1, 1]$; $(1, 3]$; $(3, \infty)$

Lösung: C, weil wir dabei mehr als 10 Beobachtungen in jeder Kategorie erwarten und weil die anderen beiden Antworten, bei denen dies auch so ist (A und B) weniger Kategorien haben.

Aufgabe 5.2

Der Apfelgroßhändler „Apple-Tree“ veröffentlich jedes Jahr eine Häufigkeitstabelle über die Gewichte einer Stichprobe seiner Äpfel, so auch letztes Jahr:

Intervall für Apfelgewicht	Anzahl
[0,100)	12
[100,200)	34
[200,300)	167
[300,400)	79
[400,∞)	65

Mit welchem Test können Sie prüfen, ob die Verteilung des Apfelgewichts mit einer Normalverteilung verträglich ist?

A: Shapiro-Wilk-Test

B: Normalscores-Test

C: Lilliefors-Test

D: χ^2-Anpassungstest

E: Exakter Test von Fisher

F: Kruskall-Wallis-Test

Lösung: D. Da die Rohdaten nicht vorliegen, können die Tests A-C nicht verwendet werden. E und F passen überhaupt nicht aufs Problem.

Aufgabe 5.3

Der Apfelgroßhändler „Apfel-Fred“ bezieht seine Äpfel von zwei Bauern. Der zweite Bauer behauptet, dass seine Äpfel im Schnitt mehr wiegen. Um herauszufinden, ob er recht hat, wiegt Fred 95 Äpfel des ersten Bauers und 75 Äpfel den zweiten Bauers. Freds Sohn, ein Mathe-Student, fand mit einem Shapiro-Wilk-Test heraus, dass die Apfelgewichte der beiden Bauern jeweils sehr gut mit einer Normalverteilung harmonieren.

Wie kann Fred prüfen, ob die Äpfel des zweiten Bauer im Schnitt signifikant mehr wiegen?

A: Kruskall-Wallis-Tests

B: Approximativer Zweistichproben-t-Test

C: Einstichproben-t-Test

D: χ^2-Anpassungstest

E: Wilcoxon-Vorzeichen-Rang-Test

F: Wilcoxon-Rangsummentest

Lösung: B, da zwei Stichproben mit normalverteilten Daten mit evtl. verschiedener Varianz vorliegen.

Aufgabe 5.4

Sie fahren mit der Bahn zur Hochschule und notieren sich jeden Tag, wie viele Minuten Verspätung die Bahn hatte. Nach 77 Tagen sieht Ihre Tabelle so aus:

Verspätung	Anzahl Tage
[0,1)	29
[1,2)	20
[2,3)	17
[3,∞)	11

Sie möchten wissen, ob die Daten mit einer Exponentialverteilung verträglich sind und schätzen dazu zunächst λ mit einer Maximum-Likelihoodschätzung aus der Häufigkeitstabelle auf $\lambda = 0.4$.

Weil Sie der Asymptotik misstrauen, berechnen Sie den p-Wert zur entsprechenden Chi-Quadrat-Prüfgröße per Simulation:

```
#
chisq.test( c(29,20,17,11), p=c(pexp(1,0.4), pexp(2,0.4)-pexp(1,0.4),
                                pexp(3,0.4)-pexp(2,0.4), 1-pexp(3,0.4)),
            simulate.p.value = TRUE, B = 10000 )


##
##  Chi-squared test for given probabilities with simulated p-value (based
##  on 10000 replicates)
##
## data:  c(29, 20, 17, 11)
## X-squared = 10.19, df = NA, p-value = 0.0161

#
```

Was wissen Sie über den wahren p-Wert des ursprünglichen Testproblems?

A: Die Simulation irrt nicht, der p-Wert beträgt ungefähr 1.6

B: Der p-Wert ist tatsächlich größer, da ein Parameter geschätzt wurde.

C: Der p-Wert ist tatsächlich kleiner, da ein Parameter geschätzt wurde.

Lösung: C, der p-Wert beträgt ungefähr `1-pchisq(10.19,2)`$\approx 0.6\%$. Wenn man es genauer wissen will, müsste man den das mit einer selbst programmierten Simulation überprüfen, in der der Parameter zunächst aus einer Häufigkeitstabelle wie der obigen geschätzt wird.

ABC-Spiel (6)

Aufgabe 6.1

Der Apfelgroßhändler „Apfel-Fred“ bezieht seine Äpfel von vier Bauern. Die Äpfel sind in drei Kategorien eingeteilt: süß, sauer und mild. Mit welchem Test kann Fred überprüfen, ob die Verteilungen auf die drei Kategorien bei den vier Bauern voneinander abweichen?

A: Kruskall-Wallis-Tests
B: Approximativer Zweistichproben-t-Test
C: χ^2-Homogenitätstest
D: χ^2-Anpassungstest
E: Wilcoxon-Vorzeichen-Rang-Test
F: Binomialtest

Lösung: C, da drei Stichproben vorliegen und auf die Gleichheit von Multinomialverteilungen getestet werden soll.

Aufgabe 6.2

Die Wirkung eines neues Medikaments gegen Covid-19 soll auf seine Abhängigkeit vom Alter überprüft werden. Dazu wird für 400 Covid-19-Patienten zehn unterschiedlicher Alterskategorien überprüft, ob das Medikament wirkt oder nicht. Um die Unabhängigkeit zu prüfen, verwenden Sie einen χ^2-Unabhängigkeitstest. Wie viele Freiheitsgrade besitzt die Verteilung der Prüfgröße unter der Nullhypothese?

A: 8 B: 9 C: 10 D: 18 E: 20

Lösung: B, da $(10-1)\cdot(2-1)=9$.

Aufgabe 6.3

Nehmen Sie an, die Körpergröße und das logarithmierte Einkommen kann mit einer bivariaten Normalverteilung beschrieben werden. Sie wollen herausfinden, ob es einen linearen Zusammenhang zwischen der Körpergröße und dem logarithmierten Einkommen gibt. Welchen Test verwenden Sie?

A: Gauß-Test
B: Korrelationstest
C: χ^2-Unabhängigkeitstest
D: χ^2-Anpassungstest
E: Wilcoxon-Vorzeichen-Rang-Test
F: Binomialtest

Lösung: B, da der Zusammenhang zwischen zwei metrischen Variablen, die mit einer bivariaten Normalverteilung modelliert werden können, untersucht werden soll.

Aufgabe 6.4

(i) Mit dem χ^2-Test auf stochastische Unabhängigkeit kann nachgewiesen werden, dass der Behandlungserfolg vom eingesetzten Medikament unabhängig ist.

A: wahr B: falsch C: Da müsste ich raten.

Lösung: B, da mit dem χ^2-Test auf stochastische Unabhängigkeit nur signifikant nachgewiesen werden kann, dass der Behandlungserfolg vom eingesetzten Medikament abhängig ist.

(ii) Bei einer Untersuchung des Zusammenhangs zwischen der Dosierung eines Nahrungsergänzungsmittels und dem Wohlbefinden (gemessen mit einem metrischen Score) ermittelten Sie eine Stichprobenkorrelation von 0.01. Das bedeutet, dass kein Zusammenhang zwischen der Dosierung eines Nahrungsergänzungsmittels und dem Wohlbefinden vorliegt.

A: wahr B: falsch C: Da müsste ich raten.

Lösung: B, da die Stichprobenkorrelation lediglich die lineare Korrelation misst.

ABC-Spiel (7)

Aufgabe 7.1

Betrachten Sie die Tabelle:

	A	B
1	1	2
2	3	4
3	$=A\$1 + \$A2$	

Welcher Wert wird in der Zelle *B3* angezeigt, wenn Sie die Formel aus *A3* kopieren?

A: 3 B: 5 C: 6 D: 7

Lösung: B, da die Formel beim Kopieren zu $= B\$1 + \$A2$ wird.

Aufgabe 7.2

Die Spalten in Excel werden mit $A, B, C, \ldots$ benannt. Wie geht es nach „Z“ weiter?

A: AA
B: Z1
C: A1
D: ZA
E: Es geht nicht mehr weiter.
F: Da müsste ich raten.

Lösung: A

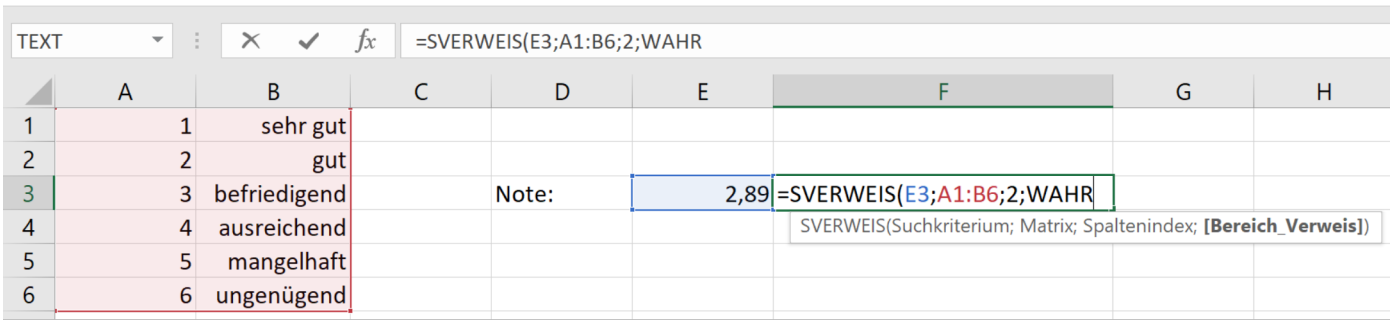
Aufgabe 7.3

Sie tippen eine Zahl in eine Zelle um, die Excel sogleich in ein Datum umwandelt. Woran liegt das?

A: Falsche Zellformatierung
B: Falsche Excel-Version
C: Gewisse Zellen sind für Datumsangaben reserviert.
D: Die Eingabe dauerte zu lange.

Lösung: A, da

Aufgabe 7.4



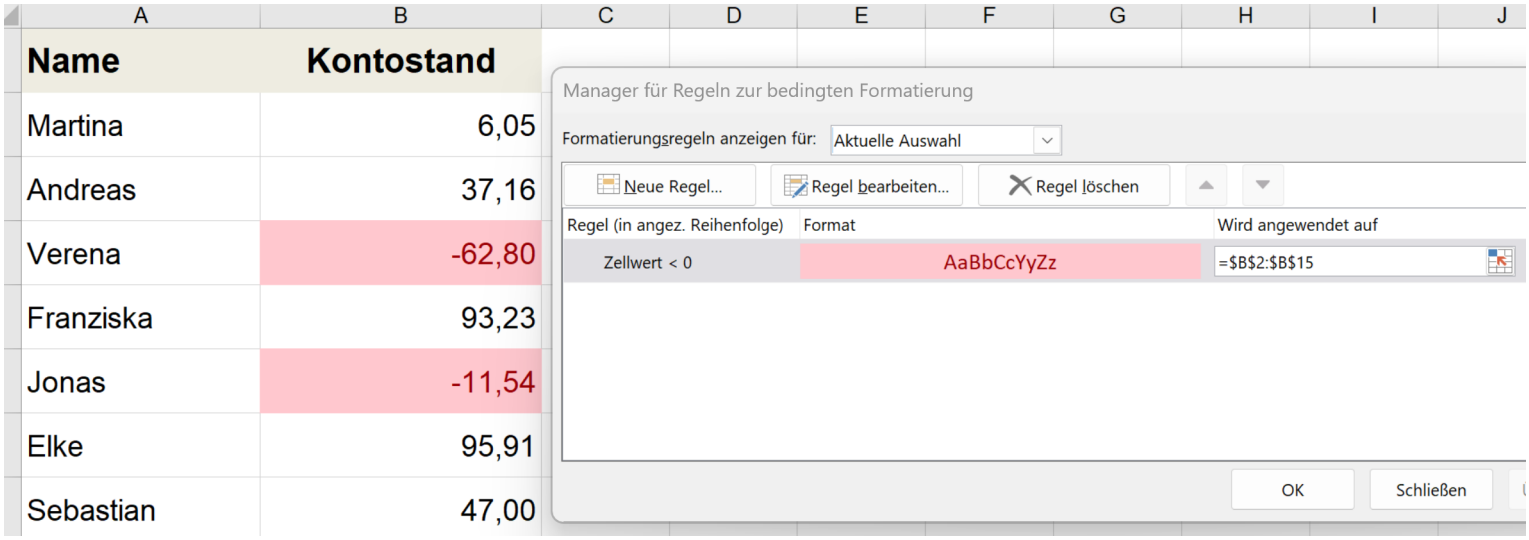
Was wird in Zelle `F3` angezeigt, wenn Sie die Formel korrekt mit „)“ abschließen?

A: 2
B: 3
C: gut
D: befriedigend
E: Fehlermeldung
F: leere Zelle

Lösung: C, da die Option `Bereich_Verweis = WAHR` den nächstkleinere Wert zu 2.89 in der ersten Spalte der Tabelle sucht (das ist 2) und dann den entsprechenden Wert in der zweiten Spalte ausgibt.

Der Tooltip ist hier sehr verwirrend. Abhilfe verspricht die neue Funktion `XVERWEIS`, die aber erst ab Excel 365 von 2019 zur Verfügung steht.

Aufgabe 7.5



Der obige Screenshot zeigt links einen kleinen Ausschnitt aus einer Tabelle, und rechts, wie die rote Hintergrundfarbe zustande kommt.

Wie färben Sie geschickt zusätzlich die Namen aller Kunden mit negativem Kontostand rot ein?

A: Ich wende die bedingte Formatierung auch auf Spalte `A` an.

B: Ich färbe die entsprechenden Felder durch Setzen der roten Hintergrundfarbe ein.

C: Ich erstelle eine bedingte Formatierung mit einer Formel zur Ermittlung der zu formatierenden Zellen.

D: Ich schreibe eine VBA-Programm.

Lösung: C, da dies bei weitem der geringste Aufwand ist und Variante A nicht funktioniert.

ABC-Spiel (8)

Aufgabe 8.1

Was müssen Sie bei `str_detect(text, Suchstring)` für `Suchstring` eingeben, um herauszufinden, ob der Ausdruck „$M \setminus N$“ in `text` vorkommt?

A: `"M\N"`

B: `"M\\N"`

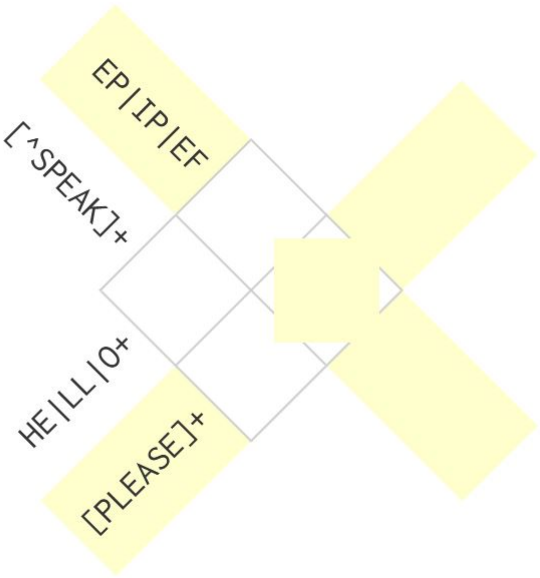
C: `"M\\\N"`

D: `"M\\\\N"`

Lösung: D, da der Backslash („\ “) sowohl im Suchstring wie auch (nochmal) im regulären Ausdruck „maskiert“ werden muss.

Aufgabe 8.2

Welcher Buchstabe gehört in das markierte Feld?



Hinweis: Die Regeln sind wie beim normalen Kreuzworträtsel: Die Einträge müssen waagrecht wie senkrecht komplett zu den regulären Ausdrücken passen, d.h. bei `str_extract()` *würde die gesamte Zeile bzw. Spalte ausgegeben.*

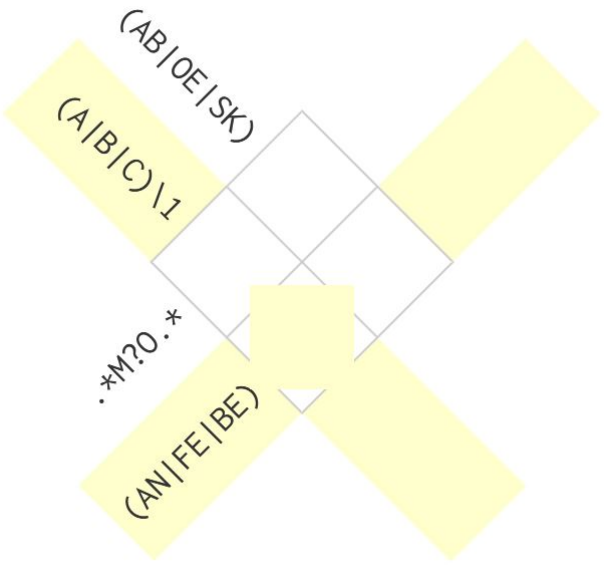
Dieses und die nächsten beiden Rätsel sind der Internetseite `https://regexcrossword.com/challenges/beginner` *entnommen. Dort gibt es noch viele weitere Rätsel dieser Art.*

A: H B: E C: L D: P E: F

Lösung: D, da „F“ nicht in der waagrechten Buchstabenmenge auftritt.

Aufgabe 8.3

Welcher Buchstabe gehört in das markierte Feld?

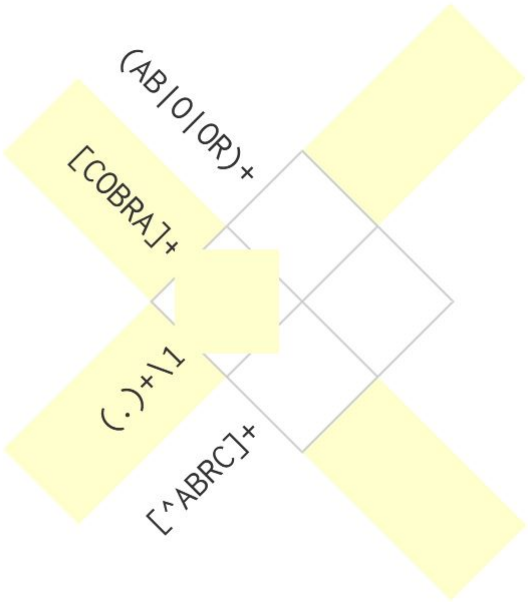


A: A B: B C: C D: D E: E F: F

Lösung: B, da wg. des senkrechten Ausdrucks nur „AA“, „BB“ und „CC“ in Frage kommen, wobei „CC“ wegen des waagrechten Ausdrucks ausscheidet. Waagrecht „AN“ ist wegen des zweiten senkrechten Ausdrucks aber nicht möglich.

Aufgabe 8.4

Welcher Buchstabe gehört in das markierte Feld?



A: C B: O C: B D: R E: A

Lösung: B, da die zweite senkrechte Spalte „OO“ sein muss (wg. des zweiten waagrechten Ausdruck). Die erste waagrechten Zeile muss daher auch „OO“ sein.

ABC-Spiel (9)

Aufgabe 9.1

Bei einem Ihrer Fragebögen gibt es eine vierstufige Ratingskala von „stimme überhaupt nicht zu“ bis „stimme voll zu“. Die Antworten, die in einem `Tibble` vorliegen, sollen nun auf die zwei Stufen „stimme nicht zu“ und „stimme zu“ komprimiert werden.

Welche Funktion bietet sich dafür an?

A:	`fct_collapse()`	C:	`fct_recode()`	E:	`fct_reorder()`
B:	`fct_lump()`	D:	`fct_infreq()`	F:	`fct_relevel()`

Lösung: `fct_collapse()`, da

```
fct_collapse( antwort,
   "stimme nicht zu", c("stimme überhaupt nicht zu", "stimme eher nicht zu"),
   "stimme zu", c("stimme voll zu", "stimme eher nicht zu") )
```

das Gewünschte liefert.

Aufgabe 9.2

Die folgende Tabelle zeigt einen Ausschnitt des Ergebnisses der Bundestagswahl 2021:

Endgültiges Wahlergebnis bei der Bundestagswahl 2021

Name der Partei	Kurzbezeichnung	Zweitstimmen	2021 in %	2
Christlich Demokratische Union Deutschlands	CDU	8.775.471	18,9	
Sozialdemokratische Partei Deutschlands	SPD	11.955.434	25,7	
Alternative für Deutschland	AfD	4.803.902	10,3	
Freie Demokratische Partei	FDP	5.319.952	11,5	
DIE LINKE	DIE LINKE	2.270.906	4,9	
BÜNDNIS 90/DIE GRÜNEN	GRÜNE	6.852.206	14,8	
Christlich-Soziale Union in Bayern e.V.	CSU	2.402.827	5,2	
Südschleswigscher Wählerverband	SSW	55.578	0,1	
FREIE WÄHLER	FREIE WÄHLER	1.127.784	2,4	
Partei für Arbeit, Rechtsstaat, Tierschutz, Elitenförderung und basisdemokratische Initiative	Die PARTEI	461.570	1,0	
PARTEI MENSCH UMWELT TIERSCHUTZ	Tierschutzpartei	675.353	1,5	
Nationaldemokratische Partei Deutschlands	NPD	64.574	0,1	
Piratenpartei Deutschland	PIRATEN	169.923	0,4	
Ökologisch-Demokratische Partei	ÖDP	112.314	0,2	
V-Partei³ - Partei für Veränderung, Vegetarier und Veganer	V-Partei³	31.884	0,1	
DEMOKRATIE IN BEWEGUNG	DiB	7.184	0,0	
Bayernpartei	BP	32.790	0,1	
Allianz für Menschenrechte, Tier- und Naturschutz	Tierschutzallianz	13.672	0,0	
Marxistisch-Leninistische Partei Deutschlands	MLPD	17.799	0,0	
Partei für Gesundheitsforschung	Gesundheitsforschung	49.349	0,1	
Menschliche Welt - für das Wohl und Glücklichsein aller	MENSCHLICHE WELT	3.786	0,0	
Deutsche Kommunistische Partei	DKP	14.925	0,0	
Die Grauen – Für alle Generationen	Die Grauen	19.443	0,0	
Bürgerrechtsbewegung Solidarität	BüSo	727	0,0	
Partei der Humanisten	Die Humanisten	47.711	0,1	
Gartenpartei	Gartenpartei	7.611	0,0	
Die Urbane. Eine HipHop Partei	du.	17.811	0,0	

Quelle: www.bundeswahlleiter.de

Die Rohdaten, also die einzelnen Stimmzettel, liegen Ihnen als R-Tibble vor. Sie wollen im Ergebnis nun nur die fünf größten Parteien auflisten, alle anderen Parteien unter „Sonstige“.

Welche Funktion ist dabei hilfreich?

A:	`fct_collapse()`	C:	`fct_recode()`	E:	`fct_reorder()`
B:	`fct_lump()`	D:	`fct_infreq()`	F:	`fct_relevel()`

Lösung: B, da

```
fct_lump( stimmzettel, n=5, other_level="Sonstige" )
```

nur die fünf am häufigsten gewählten Parteien als eigenständige Stufe erhält. Der Rest wird unter „Sonstige“ zusammengefasst.

Aufgabe 9.3

In einer Umfrage unter Akademikern wurde u.a. nach dem monatlichen Einkommen und dem Studienfach gefragt. In einer Grafik soll nun je Studienfach das mittlere Einkommen dargestellt werden.

Welche Funktion ist dabei hilfreich?

A:	`fct_collapse()`	C:	`fct_recode()`	E:	`fct_reorder()`
B:	`fct_lump()`	D:	`fct_infreq()`	F:	`fct_relevel()`

Lösung: E, da

```
fct_reorder( Studienfach, Einkommen )
```

die Studienfach-Stufen nach dem jeweiligen Median der Einkommen sortiert.

Aufgabe 9.4

Welchen Wert enthält `a[2]` nach folgender Befehlskette:

```
a <- factor(c("A","B","A")); a[2] <- "C"
```

A:	`"A"`	C:	`"C"`	E:	`2`
B:	`"B"`	D:	`NA`	F:	`3`

Lösung: D, da bei der Erzeugung von `a` automatisch die Stufen `"A"` und `"B"` angelegt wurden. Die Stufe `"C"` gibt es daher nicht.

Aufgabe 9.5

Der `str()`-Befehl liefert für den Vektor `Frage_1`, der die Antworten zu einem Item aus einer Umfrage enthält, folgende Ausgabe:
`Ord.factor w/ 6 levels "trifft überhaupt nicht zu"<..: 2 0 5 4 4 4 0 3`

D.h. das Item ist vom Typ Faktor und die Stufen werden als Zahlenwerte gespeichert.

Damit kann die mittlere Zustimmung zum Item über `mean( Frage_1 ) / 5` berechnet werden.

A: wahr B: falsch C: Da müsste ich raten.

Lösung: B, da die Werte eines Faktortyps Stufen sind und keine Zahlen. Diese müssen erst mit `as.numeric` in Zahlen umgewandelt werden, etwa `mean( as.numeric(Frage_1) ) / 5`

Aufgabe 9.6

Der `str()`-Befehl liefert für die Spalte `Frage_1` des Tibbles `dat` folgende Ausgabe:
`Ord.factor w/ 6 levels "trifft überhaupt nicht zu"<..: 2 0 5 4 4 4 0 3`

Wenn Sie nun mit `ggplot(dat) + geom_bar( aes(x=Frage_1) )` ein Säulendiagramm zeichnen lassen, werden in jedem Fall alle sechs Stufen im Diagramm zu sehen sein.

A: wahr B: falsch C: Da müsste ich raten.

Lösung: B. Wenn die Stufe „1“ nicht vorkommt, wird sie auch nicht angezeigt. Wenn man sie sehen möchte, muss man `+scale_x_discrete(drop = FALSE)` zum Befehl hinzufügen.