

D-MATL D-MAVT RW
Stochastik (401-0603-00L)

Prüfungsangabe

1. (9 Punkte) Beurteile die folgenden 18 Aussagen auf ihre Richtigkeit. Die Aussagen sind jeweils in Dreiergruppen eingeteilt. **In einer Gruppe können alle, einige oder keine einzige Aussage richtig sein. Kreuze Deine Antworten auf dem zusätzlich angehängten Antwortblatt an.** Pro korrekter Antwort gibt es einen halben Punkt. Es gibt **keinen** Punkteabzug für falsche Antworten.

a) Die Firma Tesla hat ein neues Werk in Deutschland eröffnet. Sie stellt dort spezielle Batterietypen für Elektroautos her. Die Ladedauer der Batterien sollte möglichst gleich sein wie bei den Batterien, die im Heimwerk hergestellt werden. Es wird jedoch vermutet, dass sich die Ladedauern unterscheiden. Um das zu überprüfen wird ein statistischer Test anhand zweier gleich grosser Stichproben von Batterien (je eine im Heimwerk und eine im deutschen Werk) durchgeführt.

- (i) Es handelt sich um gepaarte Stichproben.
- (ii) Bei gepaarten Stichproben sind die beiden Stichproben immer gleich gross.
- (iii) Bei einem t-Test für gepaarte Stichproben dürfen die Beobachtungen innerhalb der zwei Gruppen abhängig sein.

b) Beurteile folgende Aussagen über stetige Zufallsvariablen X, Y .

- (i) Wenn wir die gemeinsame Dichte von den zwei Zufallsvariablen (X, Y) kennen, so können wir damit immer die Kovarianz $\text{Cov}(X, Y)$ berechnen.
- (ii) Wenn X und Y je normalverteilt sind, ist $X + Y$ auch normalverteilt.
- (iii) Aus der gemeinsamen Dichte der Zufallsvariablen (X, Y) können wir nur im Falle von Unabhängigkeit die Randdichten bestimmen.

c) Sei (X, Y) eine stetige Zufallsvariable mit einer uniformen Verteilung auf dem Viereck mit Ecken $(1, 0)$, $(0, 1)$, $(-1, 0)$ und $(0, -1)$. Dann gilt:

- (i) Die Randverteilung von X ist $\text{Uniform}(-1, 1)$.
- (ii) Für (x, y) in dem Viereck gilt $f_{X,Y}(x, y) = 1/2$.
- (iii) X und Y sind unabhängig.

d) Beurteile folgende Aussagen über statistische Tests.

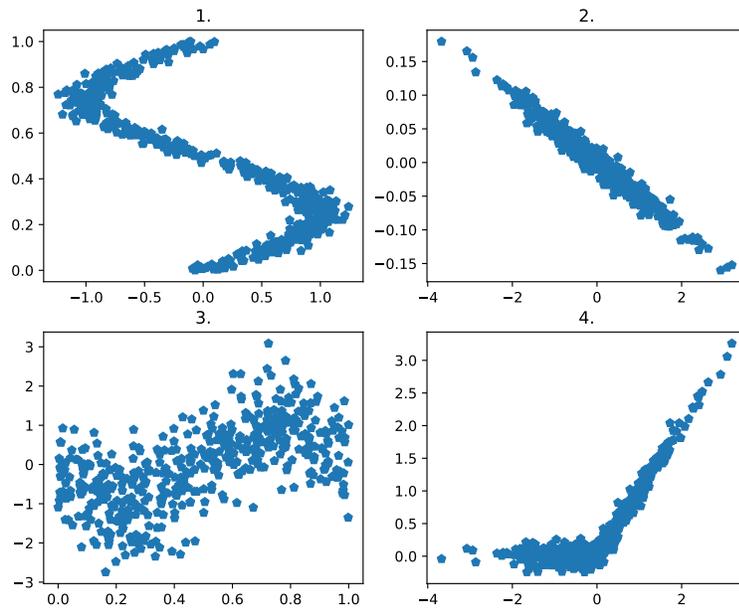
- (i) Wir betrachten einen Z-Test, bei dem die Nullhypothese auf dem 5%-Niveau nicht verworfen wird. Dann ist der p -Wert sicher kleiner als 0.05.
- (ii) Die Macht eines statistischen Tests wird grösser, je grösser das Niveau α gewählt wird.
- (iii) Die Macht eines statistischen Tests wird kleiner, je grösser der Stichprobenumfang gewählt wird.

e) Ordne die vier Scatterplots den empirischen Korrelationen

$$a = -0.98, \quad b = 0.59, \quad c = -0.78, \quad d = 0.83$$

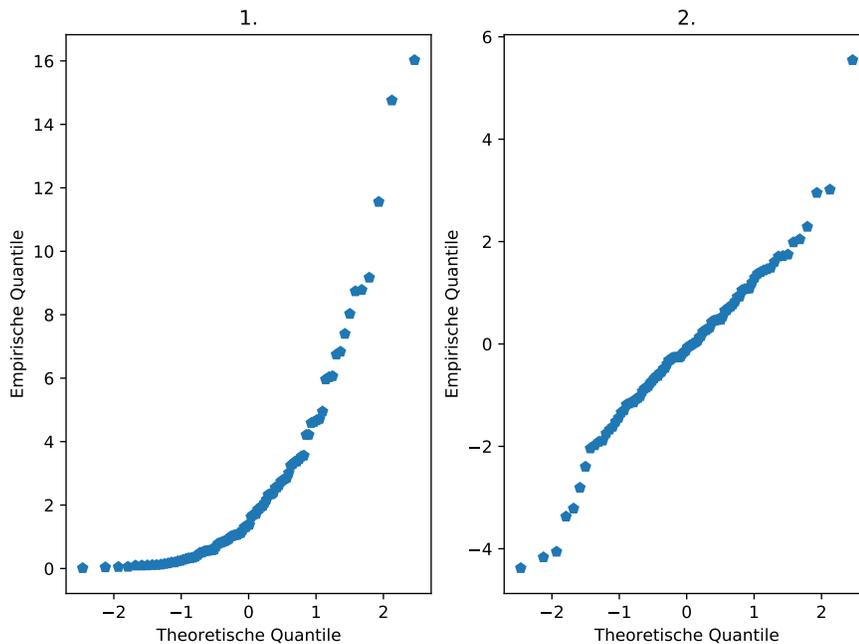
zu und beurteile die folgenden Aussagen:

- (i) Es gilt $2c, 1a$.
- (ii) Es gilt $3b, 4d$.
- (iii) Der Korrelationskoeffizient in 3. ist positiv.



f) Betrachte die folgenden QQ Normalplots von jeweils 100 Beobachtungen und beurteile die Aussagen

- (i) Die kleinste Beobachtung im 2. Plot liegt zwischen -3 und -2 .
- (ii) Die empirische Verteilung im 2. Plot hat schwerere Enden als eine Standardnormalverteilung.
- (iii) Das empirische 90%-Quantil im 1. Plot ist grösser als jenes der Standardnormalverteilung.



2. (7.5 Punkte) Wir analysieren die Chancen eines Fussballclubs, das Viertelfinale der Champions League 2021 zu erreichen, falls er es in das Achtelfinale geschafft hat. Der Club schafft es in das Viertelfinale, falls er beide Spiele (d.h. Auswärts- und Heimspiel) des Achtelfinals gewinnt. Sollte jede der beiden Mannschaften genau eines der Spiele des Achtelfinals gewinnen, so wird das Weiterkommen per Elfmeterschiessen entschieden. Mit A und H bezeichnen wir das Ereignis, dass der Club das Auswärts- bzw Heimspiel gewinnt. Einem Online-Wettbüro entnehmen wir die Wahrscheinlichkeiten

$$\mathbb{P}(A) = \frac{1}{3}, \quad \mathbb{P}(H) = \frac{2}{3}.$$

Kommt es zum Elfmeterschiessen, so gewinnt der Club es mit Wahrscheinlichkeit $1/2$. Bezeichne mit E das Ereignis, dass der Club das Elfmeterschiessen gewinnt. Ausserdem bezeichnen wir mit V das Ereignis, dass der Club das Viertelfinale erreicht. Wir nehmen an, dass A , H und E unabhängig sind

- a) (1.5 Punkte) Bestimme die Wahrscheinlichkeit, dass der Club das Viertelfinale *nach dem Elfmeterschiessen* erreicht.
Hinweis: Gesucht ist die Wahrscheinlichkeit, dass es zum Elfmeterschiessen kommt und der Club dieses gewinnt.
- b) (1.5 Punkte) Bestimme die Wahrscheinlichkeit, dass der Club das Viertelfinale erreicht.
- c) (1.5 Punkte) Das Auswärtsspiel resultiert in einem Sieg für den Club. Bestimme die Wahrscheinlichkeiten, dass der Club das Viertelfinale erreicht bzw. nicht erreicht, gegeben dieses Ergebnis der Hinrunde, d.h. bestimme $\mathbb{P}(V|A)$ und $\mathbb{P}(V^c|A)$.

Nehmen wir nun an, der Club hätte das Auswärtsspiel verloren. Die Wahrscheinlichkeit, dass er dennoch ins Viertelfinale kommt berechnet sich zu $p = \mathbb{P}(V|A^c) = 1/3$. Wir möchten überprüfen, ob $p = \mathbb{P}(V|A^c) = 1/3$ historisch belegt werden kann und entscheiden uns dazu, p eigenständig aus historischen Daten zu schätzen.

- d) (1.5 Punkte) In den vergangenen 3 Jahren schafften es $x_1 = 37.72\%$, $x_2 = 30.79\%$ und $x_3 = 20.98\%$ der Teams, die das Auswärtsspiel im Achtelfinale verloren hatten, in das Viertelfinale. Wir nehmen an, dass dieser Prozentsatz X einer Betaverteilung mit Parametern $\alpha > 0$ und $\beta > 0$ folgt, d.h. $X \sim \text{Beta}(\alpha, \beta)$. Erwartungswert und Varianz dieser Verteilung sind gegeben als

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta} \quad \text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Des weiteren nehmen wir an, dass $\beta = 2$, und möchten den verbleibenden, unbekanntem Parameter α schätzen. Bestimme den Momentenschätzer $\hat{\alpha}^{\text{MOM}}$ für α .

- e) (1.5 Punkte) Berechne die Varianz der geschätzten Verteilung Beta ($\hat{\alpha}^{\text{MOM}}, \beta$). Wir entschliessen uns dazu, den konkreten Wert $p = 1/3$ anzunehmen, wenn p innerhalb zwei Standardabweichungen des Mittelwerts der geschätzten Verteilung Beta ($\hat{\alpha}^{\text{MOM}}, \beta$) liegt. Ist dies der Fall?

3. (10 Punkte) In einem Teamwettbewerb versuchen Bonnie und Clyde möglichst viele Sekunden auf je einer Slackline zu balancieren. In jeder Sekunde ist die Wahrscheinlichkeit von der Leine zu fallen gegeben als $p \in [0, 1]$, sowohl für Bonnie als auch Clyde. Ausserdem sind die Zeiten, die Bonnie und Clyde auf der Slackline balancieren unabhängig voneinander. Bonnie und Clyde scheiden aus dem Bewerb, sobald einer der beiden von der jeweiligen Leine fällt. Sei X die Sekunde, in der Bonnie von der Leine fällt, sei Y die Sekunde, in der Clyde von der Leine fällt und sei Z die Sekunde, in der das Paar aus dem Bewerb ausscheidet.

- a) (1 Punkte) Gib die Verteilung von Y inklusive Parameter an.
- b) (2 Punkte) Gegeben, dass Clyde mindestens s volle Sekunden auf seiner Slackline balanciert, betrachte die Wahrscheinlichkeit, dass er noch mindestens t weitere volle Sekunden auf der Leine balancieren kann. Zeige, dass diese Wahrscheinlichkeit nicht von s abhängt. Wie lautet der konkrete Wert für $t = 3$ und $p = 0.5$?
- c) (2 Punkte) Zeige, dass Z geometrisch verteilt ist, d.h. $Z \sim \text{Geom}(q)$, und bestimme den Parameter q .
- d) (2 Punkte) Nehmen wir nun an, Z ist geometrisch verteilt mit Parameter q . Bestimme die Loglikelihoodfunktion für q basierend auf n i.i.d. Beobachtungen z_1, \dots, z_n und leite damit den Maximum-Likelihood Schätzer \hat{q}^{MLE} in allgemeiner Form her.
Bemerkung: Dieser Punkt lässt sich auch beantworten, wenn man Punkt c) nicht gelöst hat.
- e) (1 Punkte) Bevor sie an dem Wettbewerb teilnahmen haben sich Bonnie und Clyde 10 Mal getroffen und das Balancieren ausprobiert. Dabei haben sie die folgenden Zeiten aufgezeichnet

x	26	9	2	11	21	9	6	1	3	15
y	4	12	20	9	40	2	28	3	5	16

Berechne den konkreten Wert des Schätzers aus d) für diese Beobachtungen.

- f) (2 Punkte) Clyde behauptet, er könne im Mittel länger auf einer Slackline balancieren als Bonnie. Führe an Hand der Stichproben aus Punkt e) einen Wilcoxon Test zum 5%-Niveau durch um seine Behauptung zu überprüfen. Gib geeignete Null- und Alternativhypothesen an. Welchen Wert nimmt die realisierte Teststatistik an? Wie lautet der Testentscheid?

Hinweis: Wenn mehrere Werte den gleichen Rang bekommen, wird jedem dieser Werte der Mittelwert der entsprechenden Ränge zugeordnet.

4. (9.5 Punkte) Der Bundesrat möchte den Anteil p der Corona-positiven Personen in der Bevölkerung schätzen. Dafür wird ein Corona-Massentest mit n zufällig ausgewählten Personen durchgeführt. Der Test, der dabei verwendet wird, liefert wahrheitsgetreue Ergebnisse.

- a) (1 Punkte) Sei X_n die Anzahl derjenigen Personen, deren Testergebnis Corona-positiv ist. Bestimme die Verteilung von X_n inklusive Parameter.
- b) (3 Punkte) Innerhalb der Personengruppe der Grösse $n = 500$ wurden 15 Personen positiv getestet. Betrachte $\frac{1}{n}X_n$ als Schätzer für p und bestimme den geschätzten Standardfehler. Bestimme ein 99%-Vertrauensintervall für p mittels Normalapproximation.

Neue ökonomische und sozialpolitische Massnahmen werden implementiert, von denen man sich eine Reduktion von p auf unter 0.02 erhofft. 14 Tage nach Implementierung dieser Massnahmen soll daher versucht werden, statistisch zu belegen, dass p auf unter 0.02 gefallen ist.

- c) (3 Punkte) Betrachte die Nullhypothese $H_0 : p = 0.02$. Formuliere eine geeignete Alternativhypothese um diese Reduktion zu belegen. Bestimme mittels Normalapproximation und Transformation zur Standardnormalverteilung den zugehörigen approximativen Verwerfungsbereich für X_n auf dem Signifikanzniveau 0.05.
- d) (1 Punkte) Nach Ablauf dieser 14 Tage nach Implementierung der Massnahmen werden nochmals 500 neu ausgewählte Testpersonen auf das Coronavirus getestet. Unter diesen Coronatests fallen 4 positiv aus. Wie lautet das Resultat des statistischen Tests aus Aufgabenteil c)?
- e) (1.5 Punkte) Bestimme mittels Normalapproximation und Transformation zur Standardnormalverteilung die Macht des Tests für $p = 0.01$.

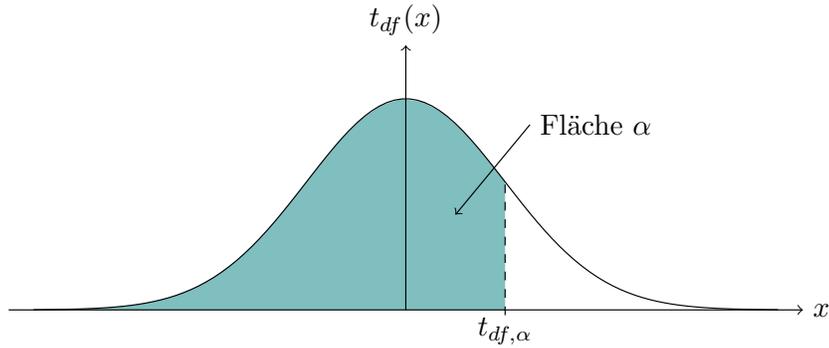
Stochastik - Tabellen

Kritische Grenzen beim Wilcoxon-Test für das 5%-Niveau:

n	zweiseitig		einseitig	
	l	u	l	u
6	0	21	2	19
7	2	26	3	25
8	3	33	5	31
9	5	40	8	37
10	8	47	10	45
11	10	56	13	53
12	13	65	17	61
13	17	74	21	70
14	21	84	25	80
15	25	95	30	90
16	29	107	35	101
17	34	119	41	112
18	40	131	47	124
19	46	144	53	137
20	52	158	60	150
21	58	173	67	164
22	65	188	75	178
23	73	203	83	193
24	81	219	91	209
25	89	236	100	225
26	98	253	110	241
27	107	271	119	259
28	116	290	130	276
29	126	309	140	295
30	137	328	151	314

Für den zweiseitigen Test ist der Verwerfungsbereich gegeben durch $K = \{W \leq l\} \cup \{W \geq u\}$.
Bei einem einseitigen Test verwendet man die entsprechenden Werte in der Spalte "einseitig".

Quantile der t-Verteilung: (df bezeichnet den Freiheitsgrad)



Lesebeispiel Tabelle: $t_{9, 0.975} = 2.262$

$df \setminus \alpha$	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576