Stochastik
(BSc D-MAVT / BSc D-MATH / BSc D-MATL)

Schreiben Sie für Aufgabe 2-4 stets alle Zwischenschritte und -rechnungen sowie Begründungen auf. Aufgabe 1 ist eine Multiple Choice Aufgabe (keine Begründungen notwendig).

Die für die Aufgaben benötigten Tabellen (Normalverteilung, Quantile der t-Verteilung, Verwerfungsbereiche für den Wilcoxon-Test) wurden mit der Prüfung ausgeteilt.

1. (10 Punkte)

Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Es gibt pro richtig beantwortete Frage 1 Punkt und pro falsche Antwort 1/2 Punkt Abzug. Minimal erhält man für die gesamte Aufgabe 0 Punkte. **Bitte benützen Sie das beiliegende Antwortblatt.**

a) Seien A und B disjunkte Ereignisse mit $\mathbb{P}(A) = a$ und $\mathbb{P}(B) = b$ für $a \in (0, 1)$ und $b \in (0, 1)$. Welche Aussage gilt **nicht**?

1. $\mathbb{P}(A \cap B) = ab$.
2. $\mathbb{P}(A \mid B) = 0$.
3. $\mathbb{P}(A \cup B) = a + b$.

b) Seien $X \sim \text{Bernoulli}(\frac{1}{2})$ und $Y \sim \text{Bernoulli}(\frac{1}{2})$ mit $\mathbb{E}(XY) = \frac{3}{8}$. Welche Aussage gilt?

1. $\mathbb{P}(X = 1, Y = 1) = \frac{1}{4}$.
2. $\mathbb{P}(X = 1, Y = 1) = \frac{3}{8}$.
3. $\mathbb{P}(X = 1, Y = 1)$ kann aus den gegebenen Informationen nicht bestimmt werden.

c) Sei N Poi(λ)-verteilt mit $\lambda > 0$. Dann ist die Varianz von $2N - \lambda$ gleich

1. 4λ.
2. 3λ.
3. 2λ.

Bitte wenden!

d) Seien $X_1, \ldots, X_9 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ und sei $\overline{X}_9 = \frac{1}{9}(X_1 + \cdots + X_9)$. Entscheiden Sie anhand der beiliegenden Tabelle, welche der folgenden Aussagen richtig ist.

1. $\mathbb{P}(X_9 - X_8 > 1) \approx 0.7603$.
2. $\mathbb{P}(X_9 + X_8 > 1) \approx 0.3085$.
3. $\mathbb{P}(\overline{X}_9 > 1) \approx 0.0013$.

e) Sei $X_1, X_2, \ldots$ eine Folge von i.i.d. $\operatorname{Exp}(\lambda)$-verteilten Zufallsvariablen mit $\lambda > 1$. Welche der folgenden Aussagen trifft zu?

1. $\frac{1}{n} \sum_{i=1}^{n} X_i$ konvergiert für $n \to \infty$ gegen $\frac{1}{\lambda}$.
2. Die Verteilung von $\frac{1}{2} \sum_{i=1}^{2} X_i$ lässt sich sehr gut durch eine Normalverteilung approximieren.
3. Weder 1. noch 2. trifft zu.

f) Wir betrachten eine Stichprobe $x_1, \ldots, x_n$ bestehend aus reellen Zahlen. Welche Kennzahl(en) kann man aus der empirischen Verteilungsfunktion dieser Stichprobe bestimmen?

1. Sowohl den empirischen Mittelwert, als auch den empirischen Median.
2. Den empirischen Mittelwert, nicht aber den empirischen Median.
3. Den empirischen Median, nicht aber den empirischen Mittelwert.

g) Seien X und Y zwei Zufallsvariablen. Welche der folgenden Aussagen trifft zu?

1. Falls X und Y abhängig sind, dann ist ihre Korrelation nicht 0.
2. Falls die Korrelation zwischen X und Y nicht 0 ist, dann sind sie abhängig.
3. Falls $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, dann sind X und Y unabhängig.

h) Welche der folgenden Aussagen trifft zu?

1. Je grösser das Signifikanzniveau eines statistischen Tests, desto kleiner ist sein Verwerfungsbereich.
2. Wenn der Wert der Teststatistik nicht im Verwerfungsbereich liegt, kann ein Fehler 1. Art nicht auftreten.
3. Tritt ein Fehler 2. Art auf, dann liegt der Wert der Teststatistik im Verwerfungsbereich.

i) Seien $a < b$ reelle Zahlen. Wir nehmen an $I = [a, b]$ sei das 95%-Vertrauensintervall für den Mittelwert von normalverteilten Daten, das mit Hilfe eines (zweiseitigen) t-Tests aus einer Stichprobe vom Umfang $n = 50$ ermittelt wurde. Wie muss der Stichprobenumfang gewählt werden, um die Länge $b - a$ des Vertrauensintervalls zu halbieren?

1. 100 Datenpunkte.
2. 200 Datenpunkte.
3. Die Länge des Vertrauensintervalls hängt von den einzelnen Werten der Stichprobe ab, und kann daher nicht allein durch den Stichprobenumfang bestimmt werden.

j) Wir wenden eine Radarpistole (ein Messgerät für die Geschwindigkeit von vorbeifahrenden Fahrzeugen) 8 Mal auf ein vorbeifahrendes Auto an. Dabei ergeben sich folgende Messwerte:

Tatsächliche km/h	10	20	30	50	70	80	100	120
Gemessene km/h	10.1	20.3	28.8	49.8	69.3	79.4	99.6	119.5

Wir fassen die beobachtete Messfehler auf als i.i.d. Realisierungen, und betrachten die Nullhypothese "Der Messfehler hat eine symmetrische Verteilung um 0" um zu entscheiden, ob die Radarpistole neu geeicht werden muss. Welche der folgenden Aussagen trifft zu?

1. Der zweiseitige Wilcoxon-Test ist anwendbar und ergibt, dass die Nullhypothese auf dem 5%-Niveau verworfen wird.
2. Der zweiseitige Wilcoxon-Test ist anwendbar und ergibt, dass die Nullhypothese auf dem 5%-Niveau nicht verworfen wird.
3. Der Wilcoxon-Test ist nicht anwendbar, da der empirische Median der Messfehler nicht 0 ist.

Bitte wenden!

2. **(8 Punkte)** Die Reagenzmethode zum Nachweis von Nitraten in natürlichen Gewässern funktioniert wie folgt: Man nimmt eine Wasserprobe und vermischt sie mit der sogenannten Reagenzlösung. Wenn die Reagenzlösung mit einer nitrathaltigen Wasserprobe vermischt wird, wird diese in 95% der Fälle rot verfärbt. Wenn die Reagenzlösung mit einer nitratfreien Wasserprobe vermischt wird, wird diese in 10% der Fälle dennoch rot verfärbt, da auch andere Chemikalien in der Wasserprobe vorhanden sein können, die auf diese Lösung so reagieren wie Nitrate. Erfahrungsgemäss sind 30% der ins Labor gesandten Wasserproben nitrathaltig.

a) Man wählt eine der ins Labor gesandten Wasserproben zufällig aus, und testet diese mit der Reagenzmethode. Berechnen Sie die Wahrscheinlichkeit, dass sie sich rot verfärben wird.

b) Angenommen man wählt eine der ins Labor gesandten Wasserproben zufällig aus, testet diese mit der Reagenzmethode und die Probe verfärbt sich rot. Berechnen Sie die Wahrscheinlichkeit, dass die getestete Wasserprobe nitrathaltig war.

Sei nun $p \in [0, 1]$ ein Parameter.

c) Es wurden $100p\%$ der Wasserproben im Labor zufällig (und insbesondere unabhängig vom Nitratgehalt) ausgewählt, die vom neuen Praktikanten des Labors präpariert werden sollten. Der Praktikant gibt versehentlich eine Nitratlösung in alle von ihm präparierten Wasserproben. Wie gross ist die Wahrscheinlichkeit in Abängigkeit von p, dass eine im Labor zufällig ausgewählte Wasserprobe nitrathaltig ist, nachdem der Praktikant die von ihm präparierten Wasserproben zu den anderen zurückgelegt hat?

d) Der Fehler des Praktikanten fällt seinen Kollegen nur dann auf, falls die Reagenzmethode bei mindestens 50% der getesteten Wasserproben positiv ausfällt. Wie gross darf p sein, damit die Wahrscheinlichkeit eines positiven Tests höchstens 50% ist?

Siehe nächstes Blatt!

3. (9 Punkte) Andreas ist unschlüssig, ob er einen Angel-Tagesschein (Preis: 40 CHF, gültig jeweils von 6:00 Uhr bis 18:00 Uhr) für eine vielversprechende Angelbucht kaufen soll.

a) Er möchte nur dann den Angel-Tagesschein erwerben, falls er mit mindestens 90% Wahrscheinlichkeit den ersten Fisch nach spätestens 4 Stunden schon an der Angel hat. Auf der Homepage der Angelstelle findet er einen Link zu einer populärwissenschaftlichen Internetseite. Nach den dortigen Angaben ist wegen den Naturgegebenheiten dieser Bucht die Wartezeit bis der erste Fisch anbeisst, unabhängig von der Tageszeit und von der Wahl des Tages, $\mathrm{Exp}(\frac{2}{3})$-verteilt. Berechnen Sie, ob Andreas den Angel-Tagesschein erwerben soll.

b) Drei Freunde von Andreas, die (unabhängig voneinander) jeweils einmal in dieser Bucht angelten, berichteten die folgenden Wartezeiten bis der erste Fisch anbeisst:

Namen	Basti	Chris	Daniel
Wartezeit (in Stunden)	1.90	1.17	1.45

Andreas modelliert die Wartezeiten durch die Exponentialverteilung $\mathrm{Exp}(\lambda)$ mit unbekanntem Parameter $\lambda > 0$. Leiten Sie den Maximum-Likelihood Schätzer für λ her und werten Sie diesen aus für die angegebene Stichprobe.

Die Angelstelle wirbt damit, dass sie die in der Bucht geangelten Fische zu einem guten Preis (7 CHF pro Fisch) zurückkaufen. Sie behaupten, dass man den Preis des Angel-Tagesscheins leicht durch den Verkauf der geangelten Fische zurückverdienen kann.

c) Um diese Behauptung zu unterstützen, führen sie die Tagesbeute ihrer ersten sieben Kunden dieser Angelsaison an, die von 6:00-18:00 geangelt haben:

Angler	1	2	3	4	5	6	7
Anzahl Fische	10	11	7	7	6	8	9

Nehmen Sie an, dass die Anzahl N der gefangenen Fische $\mathrm{Poi}(\theta)$-verteilt ist. Schätzen Sie θ mit der Momentenmethode aus der gegebenen Stichprobe.

d) Andreas möchte nicht den ganzen Tag mit Angeln verbringen. Er modelliert die Anzahl gefangener Fische in t Stunden als $\mathrm{Poi}(\lambda t)$, wobei er $\lambda = \frac{2}{3}$ der Internetseite entnimmt. Wie lange müsste Andreas mindestens in der Bucht bleiben, damit er sich den Angel-Tagesschein mit der bis dahin erwarteten Anzahl gefangener Fische finanzieren kann?

4. (9 Punkte) Am kommenden Samstag findet bei der alpinen Ski WM 2013 in Schladming die Abfahrt der Herren statt. Pro Nation dürfen jeweils 4 Läufer starten. Der Trainer der Schweizer Abfahrtsmannschaft hat bereits 3 Athleten fix nominiert. Für den letzten noch zu vergebenden Startplatz kann er sich nicht zwischen Carlo J. oder Patrick K. entscheiden. In weiser Voraussicht hatte er bereits zu Beginn der Saison angekündigt, dass in solchen Fällen getestet werden soll, ob die bis zur WM erreichten Saisonresultate statistisch signifikant unterschiedlich sind oder nicht. Falls ja, so darf der entsprechende Läufer (für den die Daten sprechen) starten, sonst entscheidet eine faire Münze. Die von den beiden Läufern in den bisherigen 8 Saisonabfahrten erzielten Zeiten (in Minuten:Sekunden) sind dabei wie folgt:

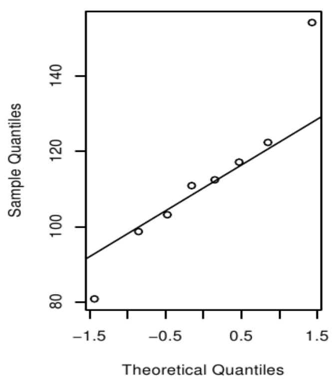
Abfahrt Nr. (i)	1	2	3	4	5	6	7	8
Carlo (a_i)	1:52.43	2:02.37	1:20.78	1:38.67	2:34.23	1:57.13	1:43.14	1:50.88
Patrick (b_i)	1:53.20	2:03.40	1:20.66	1:39.13	2:33.41	1:58.58	1:42.82	1:51.04

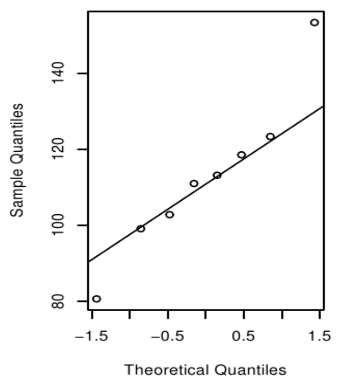
Wenn wir die Stichprobendifferenz mit $d_i := a_i - b_i$ $(i = 1, \dots, 8)$ bezeichnen, dann ergeben sich daraus die folgenden empirischen Kennwerte (in Sekunden): Mittelwert: $\overline{a}_8 = 112.45$, $\overline{b}_8 = 112.78$, $\overline{d}_8 = -0.33$. Streuung: $s_a = 21.21$, $s_b = 21.1$, $s_d = 0.75$. Gepoolte Streuung von a und b: $s_{pool} = 21.16$.

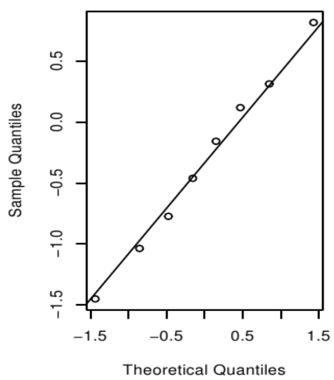
Es soll nun ein entsprechender Test entwickelt werden um zu prüfen, ob einer der beiden Skifahrer signifikant schneller ist. Und zwar so, dass für jeden einzelnen der beiden Kontrahenten, unter der Annahme er sei gleich schnell wie sein Konkurrent, die Wahrscheinlichkeit dass bereits durch den Test (also ohne dass es zu einem fairen Münzwurf kommt) gegen ihn entschieden wird, maximal 10% beträgt.

a) Ist die Stichprobe gepaart oder ungepaart? Ist der Test ein- oder zweiseitig? Wie lautet das Niveau α? Mit welcher Wahrscheinlichkeit kommt es zu einem Münzwurf wenn beide Sportler gleich schnell sind?

b) Die folgenden Graphiken sind QQ-Plots der Stichproben $(a_1, \dots, a_8)$, $(b_1, \dots, b_8)$ sowie $(d_1, \dots, d_8)$ (in Sekunden).







Siehe nächstes Blatt!

Ist die Annahme der Normalverteilung für die für den Test relevanten Daten gerechtfertigt? Geben Sie eine Begründung an. Welcher Test ist daher (unter den üblichen i.i.d. Annahmen) am besten geeignet?

c) Entwickeln Sie den Test, d.h. geben Sie Null- und Alternativhypothese, die Teststatistik sowie den Verwerfungsbereich an.

d) Für welche Werte der Teststatistik würde (ohne dass es zu einem Münzwurf kommt) gegen Carlo entschieden werden?

e) Führen Sie den Test durch. Wie entscheidet der Test? Kommt es zu einem Münzwurf oder nicht?