

## Stochastik

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

- **Bitte ...**
  - Lege deine Legi auf den Tisch.
  - Trage deine Daten in dieses Deckblatt ein, und schreibe auf jedes Blatt deinen Namen.
  - Schreibe alle **Zwischenschritte und -rechnungen** sowie **Begründungen** auf.
  - Beginne jede Aufgabe auf einem neuen Blatt.
- Es dürfen sich nur erlaubte **Hilfsmittel** auf dem Tisch befinden, d.h. 10 hand- oder maschinengeschriebene A4-Seiten und Taschenrechner ohne Kommunikationsmöglichkeiten.
- Die für die Aufgaben benötigten **Tabellen** (Verwerfungsbereiche für den Wilcoxon-Test, Normalverteilung, Quantile der t-Verteilung) wurden mit der Prüfung ausgeteilt.
- **Viel Erfolg!**

<b>Name:</b>	
<b>Vorname:</b>	
Stud. Nr.:	

Das Folgende bitte nicht ausfüllen!

Aufgabe	mögliche Punkte	erreichte Punkte	Kontrolle
1	6		
2	6		
3	6		
4	6		
5	6		

Punktetotal:	
Vollständigkeit:	



1. (6 Punkte) Eine Bekleidungsfabrik will für die nächsten 5 Jahre neue Nähmaschinen von einem Schweizer Nähmaschinenhersteller kaufen. Sie untersucht die Kosten. Der früheren Statistik zufolge liegt die Wahrscheinlichkeit bei 30%, dass bei einer Nähmaschine innerhalb dieser 5 Jahre ein kleiner Defekt anfällt. Für eine solche Maschine liegt die Wahrscheinlichkeit, dass sie dann innerhalb der 5 Jahre komplett ausfällt, bei 75%. Diese Maschine wird dann *unbrauchbar*, und man muss sie austauschen. Wenn eine Maschine keinen kleinen Defekt hat, so hat sie eine Wahrscheinlichkeit von 40%, dass sie (direkt) komplett *unbrauchbar* wird.

- a) (1.5 Punkte) Wie gross ist die Wahrscheinlichkeit, dass eine Nähmaschine, die mehr als 5 Jahre alt ist, unbrauchbar ist?
- b) (1.5 Punkte) Wie gross ist die Wahrscheinlichkeit, dass eine Nähmaschine, die unbrauchbar ist, nie einen Defekt gehabt hat?

Die Bekleidungsfabrik will nun 10 neue Nähmaschinen kaufen.

- c) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass keine dieser Nähmaschinen innerhalb von 5 Jahren einen kleinen Defekt erleidet (unter der Annahme, dass die Nähmaschinen unabhängig voneinander sind)?
- d) (1 Punkt) Der Verkaufspreis einer Nähmaschine ist 600 Franken, inklusive die Reparaturkosten. Zusätzlich zu den 10 neuen Nähmaschinen wird im Laufe der 5 Jahre jede Maschine, die komplett ausfällt, einmal durch eine neue ersetzt. Wie hoch sind *im Durchschnitt* die Gesamtkosten für die nächsten 5 Jahre?
- e) (1 Punkt) Die Bekleidungsfabrik kann nicht mehr als 6600 Franken für Nähmaschinen ausgeben. Wie gross ist die Wahrscheinlichkeit, dass dieses Budget überschritten wird?

2. (6 Punkte) Ein italienischer Küchenchef macht Pizzen zufällig. Er ist nicht präzise mit den Mengen von Mozzarella und Tomatensauce, die er verwendet. Er hat aber zwei Regeln: er benutzt nie mehr als 100 g Mozzarella, und er verwendet immer weniger Tomatensauce als Mozzarella. Sei  $X$  die Menge von Mozzarella und  $Y$  die Menge von Tomatensauce für eine Pizza, beide in hg (1 hg = 100 g). Die gemeinsame Dichte von  $X$  und  $Y$  sei gegeben durch

$$f_{X,Y}(x, y) = 8xy \mathbf{1}_{\{0 \leq x \leq 1\}} \mathbf{1}_{\{0 \leq y \leq x\}}.$$

- a) (1 Punkt) Zeigen Sie, dass  $f_X(x) = 4x^3 \mathbf{1}_{\{0 \leq x \leq 1\}}$ .
- b) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass der Küchenchef weniger als 50 g Mozzarella für eine Pizza verwendet?
- c) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass der Küchenchef mehr als  $\delta$  mal soviel Mozzarella wie Tomatensauce für eine Pizza verwendet, wobei  $\delta > 1$ ? Welcher Quotient  $\delta$  zwischen Mozzarella und Tomatensauce für eine Pizza wird mit einer Wahrscheinlichkeit von  $\frac{1}{9}$  überschritten?
- d) (1 Punkt) Wie gross ist der erwartete Quotient zwischen Mozzarella und Tomatensauce für eine Pizza?
- e) (1 Punkt) Wenn der Küchenchef schon 70 g Mozzarella für eine Pizza verwendet hat, was ist dann die erwartete Menge an Tomatensauce, die er benutzen wird?

f) (1 Punkt) Sind  $X$  und  $Y$  unabhängig? Begründen Sie Ihre Antwort durch ein präzises mathematisches Argument.

3. (6 Punkte) Die Regierung von Frankreich will die Lebensdauer ihrer  $N$  Kernkraftwerke ermitteln. Sie modellieren die Lebensdauer eines Kernkraftwerks mit der Zufallsvariable  $X$  (in Jahren). Sie benutzen für  $X$  die Gamma( $\alpha, \lambda$ )-Verteilung mit der Dichte

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} 1_{\{x \geq 0\}},$$

wobei  $\alpha > 0$  and  $\lambda > 0$  Parameter sind und  $\Gamma(\alpha)$  eine durch  $\alpha$  festgelegte Konstante. Der Erwartungswert und die Varianz einer Zufallsvariable mit einer Gamma( $\alpha, \lambda$ )-Verteilung sind  $E(X) = \frac{\alpha}{\lambda}$  und  $\text{Var}(X) = \frac{\alpha}{\lambda^2}$ . Die Regierung will die Parameter  $\alpha$  und  $\lambda$  aus einer Stichprobe mit  $n$  Beobachtungen schätzen.

a) (1.5 Punkte) Bestimmen Sie die Momentenschätzer für  $\alpha$  und  $\lambda$ . Schreiben Sie die allgemeine Formel für  $n$  Beobachtungen hin.

b) (1.5 Punkte) Nehmen Sie nun an, dass der Parameter  $\alpha$  bekannt ist. Bestimmen Sie den Maximum-Likelihood-Schätzer für  $\lambda$ . Schreiben Sie die allgemeine Formel für  $n$  Beobachtungen hin.

Für  $i = 1, \dots, N$  sei  $X_i$  jene Zufallsvariable, welche die Lebensdauer des  $i$ -ten Kernkraftwerks repräsentiert. Sei  $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$  die Zufallsvariable, welche die durchschnittliche Lebensdauer der  $N$  Kernkraftwerke darstellt.

c) (1.5 Punkte) Wie gross ist etwa die Wahrscheinlichkeit, dass die durchschnittliche Lebensdauer der  $N$  Kernkraftwerke eine Höchstgrenze  $A$  überschreitet? Treffen Sie dabei geeignete Annahmen, und beschreiben Sie diese. Drücken Sie diese Wahrscheinlichkeit mit  $\alpha, \lambda, N, A$  und der kumulativen Verteilungsfunktion  $\Phi$  der Standardnormalverteilung aus.

d) (1.5 Punkte) Wie gross muss  $N$  approximativ sein, damit die Abweichung der durchschnittlichen Lebensdauer der  $N$  Kernkraftwerke von der erwarteten Lebensdauer eines Kernkraftwerks mit einer Wahrscheinlichkeit von mindestens 95% höchstens 1 Jahr beträgt? Geben Sie das Ergebnis in Abhängigkeit von  $\alpha, \lambda$  an.

4. (6 Punkte) Ein grosser Computerhersteller bezieht Chips von einem chinesischen Chiphersteller. Er stellt fest, dass im Durchschnitt ungefähr 3% der Chips nicht funktionieren. Laut ihrem Vertrag sollten im Durchschnitt nicht mehr als 2.5% der Chips einen Ausfall haben. Der Chiphersteller denkt, dass er den Vertrag einhält. Um diese Hypothese nachzuprüfen, hat der Chiphersteller in 10 Paketen mit jeweils 1000 Chips gezählt, wieviele fehlerhaft sind, und folgende Werte erhalten.

$k$	1	2	3	4	5	6	7	8	9	10
$x_{(k)}$	22	25	26	27	28	30	31	32	33	36

Die zugehörigen empirischen Werte sind  $\bar{x}_{10} = 29$  und  $s_{10} = 4.19$ . Ausserdem nehmen wir an, dass diese Daten aus einer  $\mathcal{N}(\mu, \sigma^2)$ -verteilten Stichprobe stammen.

- a) (1 Punkt) Konstruieren Sie das 95%-Vertrauensintervall (zweiseitig) für den Parameter  $\mu$ .
- b) (1 Punkt) Lässt sich auf dem Niveau von 1% mit obigen Daten tatsächlich nachweisen, dass mehr als 2.5% der Chips fehlerhaft sind? Formulieren Sie dazu geeignete Hypothesen, geben Sie an, ob der Test ein- oder zweiseitig ist, und führen Sie den Test durch.
- c) (1 Punkt) Geben Sie das kleinste Niveau an, bei dem der Test aus b) die Nullhypothese (gerade noch) verwirft. Wie heisst dieses kritische Niveau?

Wir idealisieren nun die Situation und nehmen an, dass  $\sigma = 4.2$  bekannt ist.

- d) (1 Punkt) Bestimmen Sie den Verwerfungsbereich des analogen Tests für die Frage aus b). Wie lautet nun die Teststatistik, und wie entscheidet dieser Test?
- e) (1 Punkt) Berechnen Sie die Wahrscheinlichkeit eines Fehlers 2. Art für den in d) bestimmten Test, wenn der wahre Wert  $\mu = 26$  ist.
- f) (1 Punkt) Wie gross muss der Stichprobenumfang  $n$  gewählt werden, damit das (zweiseitige) Vertrauensintervall zum 1%-Niveau schmaler als 2 ist? Welcher empirisch gemessene Mittelwert würde dann schon zum (einseitigen) Verwerfen der Nullhypothese führen?

5. (6 Punkte) Youtube möchte zwei Werbungen A und B innerhalb eines Videos vergleichen. Wenn die Werbung gut genug ist, warten die Leute bis zum Ende der Werbung und schauen dann das Video zu Ende. Wenn die Werbung nicht gut genug ist, stoppen sie das Video und verlassen Youtube. Youtube testet die Werbungen in 2 Wochen in jeweils 10 Videos. Jede Woche werden die Videos von 1000 zufällig ausgewählten Youtube-Nutzern angeschaut, in Woche 1 mit Werbung A und eine Woche später mit Werbung B. Sie testen, wie lange es geht (in Sekunden), bis die Youtube-Nutzer das Video stoppen. In der folgenden Tabelle stehen die durchschnittlichen Zeiten (in Sekunden).

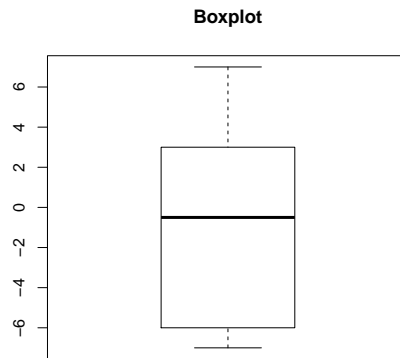
Video ( $i$ )	1	2	3	4	5	6	7	8	9	10
Werbung A ( $x_i$ )	156	155	162	160	159	165	160	164	163	168
Werbung B ( $y_i$ )	149	151	159	158	160	165	162	170	170	175
$d_i = x_i - y_i$	7	4	3	2	-1	0	-2	-6	-7	-7

Es ergeben sich daraus die folgenden empirischen Kennwerte:

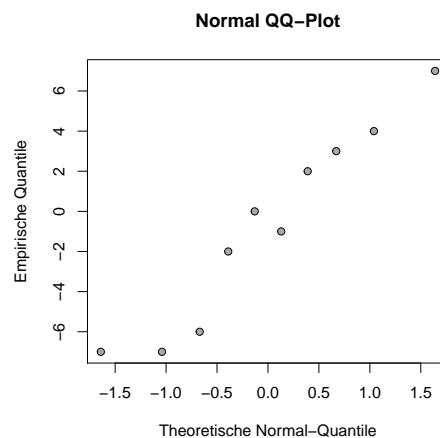
Mittelwert:  $\bar{x}_{10} = 161.20$ ,  $\bar{y}_{10} = 161.90$ ,  $\bar{d}_{10} = -0.70$ .

Streuung:  $s_x = 4.02$ ,  $s_y = 8.33$ ,  $s_d = 4.85$ .

- a) (0.5 Punkte) Ist die Stichprobe gepaart oder ungepaart? Warum?
- b) (0.5 Punkte) Youtube hat einen Boxplot der angegebenen Zeitdifferenzen  $d_i$  erstellt. Was kann man über die Verteilung sagen? Kann man sagen, ob die Verteilung normalverteilt ist? Begründen Sie.
- c) (1 Punkt) Youtube vermutet, dass die Verteilung symmetrisch ist. Welcher Test ist daher (unter den üblichen i.i.d. Annahmen) am besten geeignet, um die gegebenen Informationen bestmöglich auszunutzen? Begründen Sie.



- d) (1.5 Punkte) Führen Sie den Test durch. Lässt sich sich auf dem Niveau von 5% tatsächlich nachweisen, ob sich die durchschnittlichen Zeiten der Wiedergabe der Videos mit den beiden Werbungen unterscheiden? Formulieren Sie dazu geeignete Hypothesen, geben Sie an, ob der Test einseitig oder zweiseitig ist, und führen Sie den Test durch.



- e) (0.5 Punkte) Youtube erstellt einen QQ-Plot für die angegebenen Zeitdifferenzen  $d_i$ . Als theoretische Verteilung wird die Normalverteilung genommen. Was kann man daraus ablesen (Verteilung, Parameter)?
- f) (1 Punkte) Basierend auf e), welcher Test ist am besten geeignet? Führen Sie diesen Test durch.
- g) (1 Punkte) Berechnen Sie den  $P$ -Wert. Was können wir über die Nullhypothese sagen?

Tabelle der Standardnormalverteilung  $P(Z \leq 1.96) = 0.975$

	0	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Tabelle der  $t$ -Verteilung  $P(T_9 \leq 2.262) = 0.975$

$df \backslash \alpha$	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
$\infty$	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

Tabelle der 5%-Verwerfungsbereiche beim Wilcoxon-Test

$N$	$H_A: m \neq 0$	$H_A: m < 0$	$H_A: m > 0$
4	—	—	—
5	—	{0}	{15}
6	{0} $\cup$ {21}	[0, 2]	[19, 21]
7	[0, 2] $\cup$ [26, 28]	[0, 3]	[25, 28]
8	[0, 3] $\cup$ [33, 36]	[0, 5]	[31, 36]
9	[0, 5] $\cup$ [40, 45]	[0, 8]	[37, 45]
10	[0, 8] $\cup$ [47, 55]	[0, 10]	[45, 55]
11	[0, 10] $\cup$ [56, 66]	[0, 13]	[53, 66]
12	[0, 13] $\cup$ [65, 78]	[0, 17]	[61, 78]
13	[0, 17] $\cup$ [74, 91]	[0, 21]	[70, 91]
14	[0, 21] $\cup$ [84, 105]	[0, 25]	[80, 105]
15	[0, 25] $\cup$ [95, 120]	[0, 30]	[90, 120]