Schriftliche Prüfung (2 Stunden)

Bemerkungen:

- Alle Unterlagen (Skripte, Bücher, weitere Ausdrucke, etc.) sowie Taschenrechner ohne Kommunikationsmöglichkeit sind erlaubt.
- Mobiltelefone sind auszuschalten!
- Lesen Sie zuerst alle Aufgaben durch! Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet!
- Die nötige Tabelle befindet sich auf der hintersten Seite dieser Prüfung.
- Der Lösungsweg muss klar ersichtlich sein (Ausnahme: Multiple-Choice Aufgaben). Wenn Ihnen für einen Aufgabenteil ein Zwischenresultat fehlt, treffen Sie eine sinnvolle Annahme und markieren Sie diese deutlich.
- Aufgaben 1, 2 und 3 sind Textaufgaben. Schreiben Sie Ihre Antworten zu diesen Aufgaben auf eigene Blätter und nicht auf das Aufgabenblatt. Benutzen Sie für jede Aufgabe separate Blätter.
- Aufgaben 4 und 5 sind Multiple-Choice Aufgaben. Jede Teilaufgabe besteht aus mehreren Aussagen. Pro Teilaufgabe kann keine, eine oder mehrere Aussagen richtig sein. Die Anworten müssen direkt auf dem Aufgabenblatt angekreuzt werden und dürfen nicht auf ein separates Blatt übertragen werden. Für jede Aussage gibt es 1/2 Punkt, wenn sie korrekt markiert wird. 1/2 Punkt wird abgezogen, wenn eine Aussage falsch markiert wird. Wenn eine Aussage nicht markiert wird, gibt es keinen Punkt, es wird aber auch kein Punkt abgezogen. Keine Teilaufgabe kann mit negativer Punktzahl abgeschlossen werden. Minimal erhält man also für eine ganze Teilaufgabe 0 Punkte.

Viel Erfolg!

1. (10 Punkte) Ein Zulieferer von Metallplatten liefert 1% fehlerhafte Produkte. Das Ereignis einer intakten Platte nennen wir A, das Ereignis einer fehlerhaften Platte demnach A^c. Die verarbeitende Firma benutzt einen visuellen Test, um herauszufinden, ob eine gelieferte Platte einwandfrei ist. Der Test erfüllt laut Hersteller folgende Eigenschaften:

$$P(T_+ | A) = 0.993$$

 $P(T_- | A^c) = 0.998$

Hier bedeutet das Ereignis T_+ , dass der Test keinen Fehler feststellt ("Alles in Ordnung!") und T_- , dass der Test einen Fehler findet ("Warnung!").

- a) Berechnen Sie die Wahrscheinlichkeiten $P(T_{+})$ und $P(T_{-})$.
- b) Berechnen Sie die (bedingte) Wahrscheinlichkeit, dass eine Platte in Ordnung ist, falls der Test nichts zu beanstanden hat.
- c) Berechnen Sie die (bedingte) Wahrscheinlichkeit, dass eine Platte fehlerhaft ist, falls das Testergebnis eine Warnung ausgibt.
- d) Was ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Platte defekt ist und der Test eine entsprechende Warnung ausgibt?
- e) Die Kosten (alle in CHF) für einen Test betragen 10. Hat der Test etwas zu beanstanden (T_{-}) , kommen weitere 100 Ersatzkosten hinzu. Wird aber eine defekte Platte vom Test übersehen $(A^{c} \cap T_{+})$, kommen stattdessen 300 Ersatzkosten hinzu, da der Fehler erst später in der Produktionskette entdeckt wird. Berechnen Sie Erwartungswert und Varianz der Kosten K für eine Metallplatte.
- f) Wir führen nun einen zweiten Test U ein. Wieder bedeutet das Ereignis U_+ "Alles in Ordnung!" und U_- "Warnung!". Wir nehmen an, dass

$$P(U_{+} | A) = P(T_{+} | A) = 0.993$$

 $P(U_{-} | A^{c}) = P(T_{-} | A^{c}) = 0.998$

und dass T_{-} und U_{-} unabhängig gegeben A (und A^{c}) sind. Letzteres bedeutet:

$$P(T_{-} \cap U_{-} | A) = P(T_{-} | A) \cdot P(U_{-} | A) \quad \text{und}$$

$$P(T_{-} \cap U_{-} | A^{c}) = P(T_{-} | A^{c}) \cdot P(U_{-} | A^{c}).$$

Wie hoch ist die (bedingte) Wahrscheinlichkeit jetzt, dass eine Platte fehlerhaft ist, falls beide Testergebnisse eine Warnung ausgeben?

2. (10 Punkte) In dieser Aufgabe geht es um das Rechnen mit stetigen Verteilungen. Wir modellieren die Lebensdauer eines Bauteils (in Wochen) mit einer Verteilung mit folgender kumulativer Verteilungsfunktion

$$F(x) = \begin{cases} 0, & x \le 0\\ 1 - \exp(-(\beta x)^{\alpha}), & x > 0 \end{cases}$$

wobei $\alpha > 0$ und $\beta > 0$ Parameter sind.

- a) Bestimmen Sie die Dichte der Verteilung.
- **b)** Was ist die Wahrscheinlichkeit, dass das Bauteil länger als 4 Wochen funktioniert, wenn $\alpha = 2$ und $\beta = \frac{1}{4}$?
- c) Was ist die Wahrscheinlichkeit, dass das Bauteil innerhalb des Intervalls [5, 10] Wochen ausfällt, wenn $\alpha = 2$ und $\beta = \frac{1}{4}$?
- d) Wie gross ist der Median der Lebensdauer (für allgemeines α , β)?
- e) Bestimmen Sie die kumulative Verteilungsfunktion von $Y = (\beta X)^{\alpha}$, wobei X eine Zufallsvariable ist, die die kum. Verteilungsfunktion F hat (mit allgemeinem α und β). Wie heisst diese Verteilung?
- f) Es seien X und Y unabhängige Zufallsvariablen, wobei X kum. Verteilungsfunktion F mit $\alpha = 2$ und $\beta = \frac{1}{4}$ hat, und Y kum. Verteilungsfunktion F mit $\alpha = 3$ und $\beta = \frac{1}{6}$ hat. Berechnen Sie $P(\min\{X,Y\} \leq 4)$.

3. (8 Punkte) In dieser Aufgabe geht es um Parameterschätzung. Wir betrachten die geometrische Verteilung. Dies ist eine diskrete Verteilung mit Wahrscheinlichkeitsfunktion

$$P(X = x) = (1 - p)^{x-1}p, \quad x = 1, 2, 3, \dots$$

wobei 0 die Erfolgswahrscheinlichkeit ist. Eine geometrisch verteilte Zufallsvariable beschreibt die Anzahl der Versuche bis sich der erste Erfolg einstellt bei unabhängigen Bernoulliversuchen mit Erfolgswahrscheinlichkeit <math>p. Wir wollen den Parameter p aus einer Stichprobe schätzen.

- a) Bestimmen Sie die Likelihoodfunktion basierend auf n unabhängigen identisch verteilten Beobachtungen x_1, \ldots, x_n einer geometrisch verteilten Zufallsvariablen.
- b) Bestimmen Sie den zugehörigen Maximum Likelihood Schätzer für p.
- c) Bestimmen Sie den Momentenschätzer für p (wieder basierend auf n unabhängigen Beobachtungen x_1, \ldots, x_n). Vergleichen Sie mit dem Maximum Likelihood Schätzer.

Hinweis: $E(X) = \frac{1}{n}$.

d) Angenommen Sie haben nur eine einzige Beobachtung x einer geometrisch verteilten Zufallsvariablen. Wir können das Experiment dann auch folgendermassen interpretieren: Es wurden x unabhängige Bernoulliversuche durchgeführt und dabei wurde genau ein Erfolg beobachtet. Schreiben Sie die Likelihoodfunktion für dieses Experiment auf: Was ist der Unterschied zu der in \mathbf{a}) gefundenen Likelihoodfunktion (für n=1)? Warum erhalten Sie den gleichen Maximum Likelihood Schätzer?

4.	. (10 Punkte) Die Firma SteelWire hat ein neues Werk in China eröffnet. Sie stellt dort Stahldrähte her. Die Zugfestigkeit der Stahldrähte sollte möglichst gleich sein wie bei den Drähten, die im Heimwerk hergestellt werden. Der Chef-Ingenieur vermutet jedoch, dass sich die Zugfestigkeiten der beiden Werke unterscheiden. Er hat deshalb je 8 Drähte ausgemessen. Die Daten (in N/mm^2) sind in folgender Tabelle ersichtlich.							
	1 2 3 4 5 6	7	8					
	(6)	12.7 03.7	1605.5 1462.3					
	Kennzahlen: $\overline{x} = 1496.9$, $\overline{y} = 1462.5$, $s_x = 64.9$, $s_y = 75.1$.							
	Wir modellieren die jeweiligen Daten mit einer Normalverteilung, d.h. X_i iid. $\sim \mathcal{N}(\mu_X, \sigma^2), Y_j$ iid. $\sim \mathcal{N}(\mu_Y, \sigma^2)$.							
	a) Betrachten Sie folgende Aussagen.							
	• Es handelt sich hier um gepaarte Stichproben. Eine Versuchseinheit ist ein Stahldraht.	ichtig I	Falsch					
	• Es handelt sich hier um ungepaarte Stichproben. Die Stahldrähte aus den beiden Werken haben nichts miteinander zu tun.	\bigcirc	\bigcirc					
	• Bei gepaarten Stichproben sind die beiden Stichproben immer gleich gross.	\bigcirc	\bigcirc					
	• Bevor wir entscheiden können, ob die Stichproben gepaart sind, müssen wir das Signifikanzniveau festlegen.	\bigcirc	0					
	b) Betrachten Sie folgende Aussagen zur Nullhypothese H_0 .							
	$\bullet \ H_0: \mu_X = \mu_Y$	ichtig I	Falsch					
	\bullet H_0 : Die erwartete Zugfestigkeit ist in beiden Werken gleich gross.	\bigcirc	\bigcirc					
	\bullet H_0 : Es gibt keinen Unterschied zwischen den Zugfestigkeiten der Drähte innerhalb eines Werkes.	\bigcirc	\bigcirc					
	$\bullet \ H_0: \mu_X \neq \mu_Y$	\bigcirc	\bigcirc					
	c) Betrachten Sie folgende Aussagen zur Alternative H_A .	H_A .						
	• H_A : Es existiert ein Unterschied zwischen den erwarteten Zugfestigkeiten der beiden Werke.	ichtig I	Falsch					
	$\bullet \ H_A: \mu_X \neq \mu_Y.$	\bigcirc	\bigcirc					
	$\bullet \ H_A: \mu_X > \mu_Y.$	\bigcirc	\bigcirc					

 $\bullet \ H_A$: Das Werk in China liefert schlechtere Qualität (im Sinne von \bigcirc

tieferen Zugfestigkeiten).

d)	Sie führen nun einen zweiseitigen Zwei-Stichproben t-Test für unabhängige Stichproben durch (unabhängig davon, was Sie oben geantwortet haben). Betrachten Sie folgende Aussagen.				
	• Die Teststatistik berechnet sich als	Richtig	Falsch		
	$T = \frac{34.4}{S_{pool}\sqrt{\frac{1}{8} + \frac{1}{8}}}.$				
	• Es ist $S_{pool}^2 = \frac{1}{16} (64.9^2 + 75.1^2).$	\bigcirc	\bigcirc		
	\bullet Der Verwerfungsbereich auf dem 5% Niveau ist gegeben durch	\bigcirc	\bigcirc		
	$K = (-\infty, -2.12] \cup [2.12, \infty).$				
	• Falls der realisierte Wert der Teststatistik nicht im Verwerfungs bereich liegt, dann haben wir damit die Nullhypothese statistisc nachgewiesen.	_	\bigcirc		
e)	Wir nehmen an, dass das 95%-Vertrauensintervall für die Differ $\mu_X - \mu_Y$ gegeben sei durch $[-41,110]N/mm^2$. Betrachten Sie folge Aussagen.				
	• Der entsprechende Test verwirft die Nullhypothese $H_0: \mu_X = \mu$ auf dem 5% Niveau.	Richtig $Y \bigcirc$	Falsch		
	• Der entsprechende Test verwirft die Nullhypothese $H_0: \mu_X = \mu$ auf dem 1% Niveau.	$Y \bigcirc$	\bigcirc		
	\bullet Wenn Abweichungen ab 150 N/mm^2 als relevant gelten, dann kan hier von einem irrelevanten Unterschied ausgegangen werden.	n 🔾	\bigcirc		
	• Wenn wir den gleichen Versuch nochmals durchführen würder	n, 🔘	\bigcirc		

dann würden wir ein Vertrauensintervall mit einer sowohl anderen

Breite wie auch mit einer anderen Lage erhalten.

5. (17 Punkte)

a) Die Ereignisse A und B seien disjunkt und P(B)>0 sowie P(A)>0. Dann gilt in diesem Fall

• $P(A^c \cap B^c) + P(B) = 1 - P(A)$

Richtig Falsch

• $P(A \cap B) < P(A)$

0 0

 $\bullet \ P(A \mid B) = P(B \mid A)$

0 0

• $P(A \cup B) > P(A)$

- 0 0
- b) Betrachten Sie folgende Aussagen zu stetigen Zufallsvariablen.

• Wenn X exponential-verteilt ist, so gilt

Richtig Falsch

$$P(X \in [0,1]) > P(X \in [a, a+1])$$

für alle a > 0.

- Wenn wir die gemeinsame Dichte von zwei Zufallsvariablen (X,Y) \bigcirc kennen, so können wir damit immer die Kovarianz Cov(X,Y) berechnen.
- Aus der gemeinsamen Dichte der Zufallsvariablen (X,Y) können \bigcirc wir nur im Falle von Unabhängigkeit die Randdichten bestimmen.
- \bullet Der Erwartungswert der Zufallsvariable Xmit kumulativer Vertei- \bigcirc lungsfunktion

$$F(x) = \begin{cases} x^5 & 0 \le x \le 1\\ 0 & x < 0\\ 1 & x > 1 \end{cases}$$

ist $\frac{5}{6}$.

c) Es sei $X \sim \mathcal{N}(0, \sigma^2)$ und Y = -X. Dann gilt

• $Y \sim \mathcal{N}(0, \sigma^2)$.

Richtig Falsch

• Var(X+Y)=0.

 \cap

• Die Korrelation zwischen X und Y ist -1.

 \circ

• $Var(E[X]) = \sigma^2$.

0 0

d) Sei X eine Zufallsvariable, die die Werte 1,2,3 oder 4 annimmt. Die Entropie von X ist definiert als

$$H(X) = -\sum_{i=1}^{4} P(X = i) \cdot \log_2(P(X = i))$$

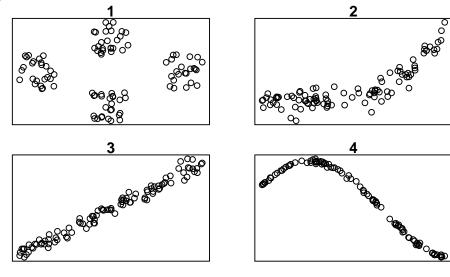
wobei $0 \cdot \log_2(0) = 0$. H(X) ist ein Mass dafür, wie "unsicher" wir über das Ergebnis von X sind.

• Es gilt immer $H(X) \ge 0$.

- Richtig Falsch
- Wenn X deterministisch ist (d.h. P(X=i)=1 für ein i), so gilt \bigcirc H(X)=0.
- Wenn H(X) = 0, so ist X deterministisch.
- Die Zufallsvariable Y = 5 X nimmt ebenfalls die Werte 1, 2, 3 und \bigcirc 4 an. Es gilt dann H(Y) = H(X).
- e) Ordnen Sie die unteren vier Scatterplots den Korrelationen

$$a = 0.01$$
 $b = 0.99$ $c = 0.81$ $d = -0.89$

zu.



• 1a, 4d

Richtig Falsch

• 2b, 3c

O C

• 1c, 2d

- \circ
- Bei keinem der Plots sind die beiden Variablen unabhängig.
- 0 0

f)	Wir modellieren die Anzahl Unfälle an einer Strassenkreuzung pro Jahr (365 Tage) mit einem homogenen Poisson-Prozess. Im Schnitt gibt es an der Kreuzung 5 Unfälle pro Jahr.	
	• Falls wir heute einen Unfall beobachtet haben, so müssen wir im O Schnitt 1/5 Jahr warten, bis der nächste Unfall geschieht.	g Falsch
	• Die Wahrscheinlichkeit, dass an einem bestimmten Tag kein Unfall Opassiert, ist gegeben durch	\bigcirc
	$e^{-5}/365$.	
	\bullet Wenn ein Unfall Kosten in der Höhe von 5000 CHF verursacht, dann \bigcirc ist die Standardabweichung der jährlichen Kosten 5000 · 5 CHF.	0
g)	Betrachten Sie folgende Aussagen zu statistischen Tests (denken Sie z.B. an einen Z -Test).	
	,	g Falsch
	\bullet Je mehr Beobachtungen wir haben, desto kleiner wird der Fehler 1. \bigcirc Art, wenn wir einen Test auf dem 5%-Niveau durchführen.	\bigcirc
	\bullet Je mehr Beobachtungen wir haben, desto wahrscheinlicher ist es , \bigcirc dass wir die Nullhypothese auf dem 5%-Niveau verwerfen, wenn sie in Tat und Wahrheit nicht stimmt.	0
h)	Wir wollen mit einem χ^2 -Test die Nullhypothese testen, ob Daten von einer Normalverteilung stammen. Dabei schätzen wir die Parameter aus der Stichprobe.	
	• Wenn wir 10 Klassen verwenden, dann müssen wir eine χ^2 - O Verteilung mit 7 Freiheitsgraden verwenden zur Ermittlung des Verwerfungsbereichs.	g Falsch
	• Die Wahl der Klassen hat keinen Einfluss auf das Ergebnis des 🔾 Tests.	\bigcirc
	• Wenn die Daten in Tat und Wahrheit von einer t-Verteilung stammen, so machen wir einen Fehler 2. Art, wenn wir die Nullhypothese nicht verwerfen.	0
	• Wenn wir die Parameter nicht aus der Stichprobe schätzen, so können wir die erwarteten Häufigkeiten auch schon vor Erhalt der Stichprobe berechnen, so lange wir wissen, wie gross die Stichprobe ist.	0

i) Lisa hat eine Maschine gekauft, die das Abwerfen einer Münze filmt und anhand einer komplizierten Videoanalyse das Ergebnis ("Kopf" oder "Zahl") vorhersagt. Der Verkäufer hat behauptet, dass man mit dieser Maschine Vorhersagen machen kann, die in mehr als der Hälfte der Fälle zutreffen.

Zu Hause möchte Lisa schauen, ob man mit Hilfe eines statistischen Tests die Behauptung des Verkäufers nachweisen kann. Sei X die Anzahl der von der Maschine korrekt geratenen Ergebnisse (aus n=20 Versuchen). Zur Modellierung von X verwenden wir also eine Binomialverteilung mit Parametern n=20 und Erfolgswahrscheinlichkeit p, d.h. $X \sim \text{Bin}(n,p), n=20$. Als Null- und Alternativhypothesen verwenden wir

$$H_0: p = 0.5 \text{ und } H_A: p > 0.5.$$

Richtig Falsch

• Falls wir 15 korrekt vorhergesagte Würfe beobachten, so berechnet O sich der p-Wert hier als

$$\sum_{k=15}^{20} {20 \choose k} 0.5^k (1 - 0.5)^{20-k}.$$

- Um die Macht zu berechnen, müssen wir nicht wissen, wie viele \cap korrekt vorhergesagte Würfe beim Versuch eingetreten sind.
- Lisa berechnet die Macht des Tests für p=0.55, der Hersteller \bigcirc für p=0.75 (beide für n=20 und für das gleiche Signifikanzniveau). Die berechnete Macht des Herstellers wird kleiner sein als die berechnete Macht von Lisa.
- Wenn wir das Signifikanzniveau verkleinern, dann wird die Macht \bigcirc generell auch kleiner.

Quantile der t-Verteilung

Bsp.: $t_{9;\ 0.975} = 2.262$

-df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
$\frac{-\alpha j}{1}$	0.325	0.727	1.376	$\frac{v_{0.90}}{3.078}$	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
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
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