Aufgabe 1 (Stichprobenverteilung, bedingter Erwartungswert) (9 + 5 = 14 Punkte)

Es seien X, Y stochastisch unabhängige Zufallsvariablen.

1. X besitze eine diskrete Verteilung mit Träger $\mathbb{N}_0$ und Dichte $f(x) = (x+1)p^2(1-p)^x, \quad x \in \mathbb{N}_0$
2. Y besitze eine diskrete Verteilung mit Träger $\mathbb{N}_0$ und Dichte $g(y) = (1-p)^y p, \quad y \in \mathbb{N}_0$

Dabei sei $p \in]0;1[$. Geben Sie in den folgenden Teilaufgaben den vollständigen Rechenweg bis zum Ergebnis an.

a) Berechnen Sie eine Dichte von $Z = X + Y$, d.h. berechnen Sie $P(Z = n)$ für $n \in \mathbb{N}_0$.

b) Berechnen Sie den bedingten Erwartungswert $E(X^Y | X = x)$ für $x \in \mathbb{N}_0, x < \frac{1}{1-p}$.
Hinweis: Sie müssen hier nicht die bedingte Dichte ausrechnen.

Aufgabe 2 (Parameterschätzung, Hypothesentests) (12 + 6 + 18 = 36 Punkte)

In der Weihnachtsmanufaktur am Nordpol wird aus abgeworfenen Rentier-Geweihen ein besonders stimmungsvoller Christbaumschmuck gefertigt. Für die drei Arbeitsschritte „Auswahl des Geweihs aus dem Lager“, „Entstauben des Geweihs“ und „Anfertigung eines Schmuckteils“ wird eine Gesamtarbeitszeit X mit einer stetigen Verteilung mit der folgenden Dichte f angesetzt:

$$f(x) = \frac{\lambda^3}{2} x^2 e^{-\lambda x} \mathbf{1}_{[0;\infty[}(x), \quad x \in \mathbb{R}$$

Dabei ist die Gesamtarbeitszeit in Minuten angegeben, $\lambda > 0$ ist unbekannt und soll im folgenden geschätzt werden, um zu klären, ob die Fertigung mit dem Bestelleingang Schritt halten kann oder ein Auftragsstau droht. In einer u.i.v.-Stichprobe $X_1, \dots, X_{70}$ von früheren Gesamtarbeitszeiten ergab sich ein Saldo von 795 Minuten.

a) Berechnen Sie eine Maximum-Likelihood-Schätzung für λ auf Grundlage der Daten.

b) Berechnen Sie eine Momentenmethode-Schätzung für λ auf Grundlage der Daten.

c) Damit es beim aktuellen Auftragseingang nicht zu einem absehbaren Auftragsstau kommt, muss die erwartete Fertigungsdauer für ein Schmuckstück weniger als 14 Minuten betragen.

(1) Prüfen Sie anhand der angegebenen Stichprobe mit einem geeigneten approximativen Test und geeigneter, in λ formulierter Hypothese und Alternative, ob Sie die Alternative zum Signifikanzniveau 5% bestätigen können. Verwenden Sie dabei die Teststatistik

$$V_n = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0}$$

mit geeigneten Werten für μ_0 und σ_0.

(2) Berechnen Sie auch den p-value des von Ihnen durchgeführten Tests und erläutern Sie anhand dieses Wertes, welche Entscheidung der Test zum 1%-Signifikanzniveau trifft.

(3) Die Gesamtarbeitszeiten mögen in einem R-Vektor `cbsdata` vorliegen. Schreiben Sie ein kurzes R-Skript, mit dem Sie anhand des Vektors den p-value aus der vorangegangenen Teilaufgabe berechnen.

Hinweise:

- In allen Teilaufgaben ist jeweils der vollständige Rechenweg anzugeben.
- In den Aufgaben b) und c) können Sie ohne weitere Rechnung folgende Identität verwenden: Für alle $n \in \mathbb{N}$ ist $\int_0^\infty x^n e^{-\lambda x} dx = \frac{n!}{\lambda^{n+1}}$.
- In Aufgabe c) können Sie die nachfolgenden Tabellen zur Normalverteilung nutzen:

Quantile der Standardnormalverteilung

$\alpha =$.900	.950	.975	.990	.995	.999	.9995
$u_\alpha =$1.28	1.64	1.96	2.33	2.58	3.09	3.29

Verteilungsfunktion der Standardnormalverteilung

	000	005	010	015	020	025	030	035	040	045	050	055	060	065	070	075	080	085	090	095
0.0	500	502	504	506	508	510	512	514	516	518	520	522	524	526	528	530	532	534	536	538
0.1	540	542	544	546	548	550	552	554	556	558	560	562	564	566	567	569	571	573	575	577

	000	005	010	015	020	025	030	035	040	045	050	055	060	065	070	075	080	085	090	095
0.2	579	581	583	585	587	589	591	593	595	597	599	601	603	604	606	608	610	612	614	616
0.3	618	620	622	624	626	627	629	631	633	635	637	639	641	642	644	646	648	650	652	654
0.4	655	657	659	661	663	665	666	668	670	672	674	675	677	679	681	683	684	686	688	690
0.5	691	693	695	697	698	700	702	704	705	707	709	711	712	714	716	717	719	721	722	724
0.6	726	727	729	731	732	734	736	737	739	741	742	744	745	747	749	750	752	753	755	756
0.7	758	760	761	763	764	766	767	769	770	772	773	775	776	778	779	781	782	784	785	787
0.8	788	790	791	792	794	795	797	798	800	801	802	804	805	806	808	809	811	812	813	815
0.9	816	817	819	820	821	823	824	825	826	828	829	830	831	833	834	835	836	838	839	840
1.0	841	843	844	845	846	847	848	850	851	852	853	854	855	857	858	859	860	861	862	863
1.1	864	865	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884
1.2	885	886	887	888	889	890	891	892	893	893	894	895	896	897	898	899	900	901	901	902
1.3	903	904	905	906	907	907	908	909	910	911	911	912	913	914	915	915	916	917	918	918
1.4	919	920	921	921	922	923	924	924	925	926	926	927	928	929	929	930	931	931	932	933
1.5	933	934	934	935	936	936	937	938	938	939	939	940	941	941	942	942	943	944	944	945
1.6	945	946	946	947	947	948	948	949	949	950	951	951	952	952	953	953	954	954	954	955
1.7	955	956	956	957	957	958	958	959	959	960	960	960	961	961	962	962	962	963	963	964
1.8	964	964	965	965	966	966	966	967	967	967	968	968	969	969	969	970	970	970	971	971
1.9	971	972	972	972	973	973	973	974	974	974	974	975	975	975	976	976	976	976	977	977
2.0	977	978	978	978	978	979	979	979	979	980	980	980	980	981	981	981	981	981	982	982
2.1	982	982	983	983	983	983	983	984	984	984	984	984	985	985	985	985	985	986	986	986
2.2	986	986	986	987	987	987	987	987	987	988	988	988	988	988	988	989	989	989	989	989
2.3	989	989	990	990	990	990	990	990	990	990	991	991	991	991	991	991	991	991	992	992
2.4	992	992	992	992	992	992	992	993	993	993	993	993	993	993	993	993	993	994	994	994
2.5	994	994	994	994	994	994	994	994	994	995	995	995	995	995	995	995	995	995	995	995
2.6	995	995	995	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996
2.7	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997
2.8	997	997	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998
2.9	998	998	998	998	998	998	998	998	998	998	998	998	998	998	999	999	999	999	999	999
3.0	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3.1	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3.2	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	1

Beispiel: $\Phi(1.240) = \Phi(1.3 + .040) \approx 0.893$. Für nicht aufgeführte x:

- $x \geq 3.2$: $\Phi(x) \approx 1$ für $x \geq 3.2$. Für $x < 0$: $\Phi(x) = 1 - \Phi(-x)$
- Interpolation: $\Phi(\lambda x + (1-\lambda)y) \approx \lambda\Phi(x) + (1-\lambda)\Phi(y)$

Aufgabe 3 (Regression) **(1 + 8 + 13 = 22 Punkte)**

Die Weihnachtsmanufaktur am Nordpol betreibt den Geschenke-Lieferservice **SANTA.nder**, mit welcher der Weihnachtsmann den Kindern die Pakete bringt. Mit einer Schlittenladung Pakete werden gleich mehrere Familien beliefert, gegebenenfalls muss die Anlieferung auch durch den Kamin erfolgen - für den Weihnachtsmann ein erhöhter Zeitaufwand im Vergleich zur Normallieferung durch die Haustür.
Lowi, der Logistikwichtel von **SANTA.nder**, hat für 8 Schlittenfuhren einmal die Anzahl der insgesamt gelieferten Pakete x und die Anzahl der durch den Kamin besuchten Familien y sowie die Gesamtlieferzeit z (in Millisekunden!) für die Fuhre in einer Tabelle erfasst und möchte wissen, ob sich die Daten durch einen Zusammenhang der Form $z = a + bx + cy$ beschreiben lassen.

Fuhre	x	y	z
1	20	4	121
2	15	3	84
3	23	6	146
4	14	8	121
5	15	7	113
6	17	2	78
7	20	10	142
8	14	5	92

Hinweis: In den folgenden Teilaufgaben ist jeweils auch der Rechenweg gesucht.

a) Erläutern Sie, welche Bedeutung a, b, c in diesem Kontext haben.

b) Berechnen Sie die Kleinste-Quadrate-Schätzung von a, b, c.

c) Die Daten der obigen Tabelle liegen in einem Data Frame `stadata`, die Attributbezeichner seien gemäß des Eingangstextes zur Aufgabe gewählt. Mit R wird die folgende Ausgabe erzeugt (hier nachträglich teilweise geschwärzt):

```
Residuals:
      1        2        3        4        5        6        7        8
 7.1660  -0.8331   5.5985   7.9184   1.9756  -9.2665 -11.1209  -1.4381

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) [redacted]   17.3962  -0.125  0.90560
[redacted]    0.9611   4.672  0.00547 **
[redacted]    1.2134   5.396  0.00295 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: [redacted] on 5 degrees of freedom
Multiple R-squared: [redacted],    Adjusted R-squared: [redacted]
F-statistic: 29.02 on 2 and 5 DF,  p-value: 0.001772
```

(1) Schreiben Sie ein kurzes R-Skript, welches diese Konsolenausgabe ergibt.
(2) Ermitteln Sie das Bestimmtheitsmaß der Regression anhand der obigen Konsolenausgabe.
(3) Berechnen Sie anhand Ihres Ergebnisses aus Teilaufgabe b), aus der Konsolenausgabe sowie der nachfolgenden Tabelle von Quantilen der t-Verteilung ein zweiseitiges 95%-Konfidenzintervall für den Parameter b.

Anhang: Quantile der $t(n)$-Verteilung

$n =$	$\alpha =$.900	.950	.975	.990	.995	.999	.9995
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	1.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65

Aufgabe 4 (Multiple Choice) (1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 12 Punkte)

Die nachfolgenden Aussagen sind entweder wahr (Antwort „ja") oder falsch bzw. nicht in jedem Fall richtig (Antwort „nein"). Geben Sie Sie jeweils die richtige Antwort an (Hinweis: Diese Aufgabe ist im Examweb-Kurs separat zu bearbeiten)

Vorlesung Datenanalyse

a)	Sind X, Y stochastisch unabhängige Poisson-verteilte Zufallsvariablen, so ist auch $Z = max(X, Y)$ Poisson-verteilt.	ja	nein
b)	Sind X, Y Poisson-verteilte Zufallsvariablen mit $P(Y = y\|X = x) = P(Y = y)$ für alle $x, y \in \mathbb{N}_0$, dann gilt auch $E(X\|Y) = E(X)$	ja	nein
c)	Sind X, Y Zufallsvariablen mit existierenden Erwartungswerten $E(XY)$ und $EY > 0$, so gilt $E(X\|Y) = E(XY)/E(Y)$.	ja	nein
d)	Ein zweiseitiger Anteilswerttest lehnt die Hypothese $H_0 : p = p_0$ ab, wenn der Wert p_0 im zweiseitigen Konfidenzintervall zur Teststichprobe liegt.	ja	nein
e)	Mit dem R-Befehl `t.test` können Zweistichprobenprobleme nur dann behandelt werden, wenn die Modellvarianzen der beiden Stichproben übereinstimmen.	ja	nein
f)	Die Durchführung einer Regression mit R terminiert nicht mit einer Fehlermeldung, wenn zwei der Regressoren im Datensatz übereinstimmen.	ja	nein
g)	Mit Ausnahme des Schätzers für die Modellvarianz sind die KQ-Parameterschätzer in der linearen Regression unter Normalverteilungsannahme auch Maximum-Likelihood-Schätzer.	ja	nein
h)	Für die Hat-Matrix $H = X(X^T X)^{-1}X^T$ gilt: $H(I - H)$ ist die Nullmatrix.	ja	nein

Vorlesung Simulation

i)	Beim M|M|1 Wartesystem stellt der Ankunftsprozess einen Poissonprozess dar.	ja	nein
j)	Das Gesetz von Little gilt grundsätzlich nicht für Wartesysteme mit einer Last-Come-First-Served (LCFS) Disziplin.	ja	nein
k)	Beim M|M|S Wartesystem entspricht die Ankunftsrate (λ) nicht der effektiven Ankunftsrate.	ja	nein
l)	In Jackson-Netzwerken können an jedem Wartesystem neue externe Kunden das Netzwerk betreten.	ja	nein

Aufgabe 5 (Zufallszahlen) (3 + 9 = 12 Punkte)

a) Definieren Sie den linearen Kongruenzgenerator formal und benennen Sie alle Elemente der Definition.

b) Wie gehen Sie vor, um einen gegebenen linearen Kongruenzgenerator unter Annahme der Unabhängigkeit der generierten Zahlen empirisch auf Uniformität zu testen? Nennen Sie ein Testverfahren und beschreiben Sie das Vorgehen beim Durchführen des Tests. (Die Verwendung von Formeln ist zur Beantwortung der Frage nicht erforderlich.)

Aufgabe 6 (Simulationsstudie) **(3 + 3 + 3 = 9 Punkte)**

a) Erläutern Sie den Unterschied zwischen Validierung und Verifikation im Rahmen einer Simulationsstudie.

b) Nennen Sie 3 Techniken zur Validierung.

c) Nennen Sie 3 Vorteile der Nutzung einer Simulationsumgebung für die Durchführung einer Simulationsstudie.

Aufgabe 7 (ARENA) **(13 + 2 = 15 Punkte)**

a) Frau M. Allwetter plant einen neuen Zoo als Touristenattraktion für die Stadt zu schaffen. Während der Planung, bittet sie Sie um Unterstützung bei der Simulation der Abläufe. Sie hofft, durch die Simulation einen besseren Eindruck über Auslastung zu erhalten und so mögliche Ressourcen-Engpässe frühzeitig identifizieren zu können. Sie beschreibt die erwarteten Abläufe wie folgt:

- Etwa alle 6 Minuten trifft ein neuer Zoobesucher ein.
- Der Besucher hat die Wahl entweder eine geführte Zooführung mit einem unserer Mitarbeitern zu machen, oder auf eigene Faust den Zoo zu erkunden. Etwa 20 % der Besucher wollen bei der Zooführung mitmachen.
- Die Teilnehmerzahl für die Führungen ist limitiert auf 20 Personen. Wir versuchen diese Teilnehmerzahl zu erreichen. Damit die Besucher allerdings nicht zu lange warten müssen, ist die maximale Wartezeit auf 40 Minuten festgelegt und die Führung wird notfalls mit geringerer Teilnehmerzahl durchgeführt.
- Die Führungen werden jeweils von einem unserer 16 Experten durchgeführt und dauern ca. 120 Minuten.
- Die Besucher, die sich gegen eine geführte Tour entscheiden, können sich ein Headset mit Audioguide mit interessanten Informationen über unsere Tiere ausleihen. Die Dauer eines Besuchs mit einem Headset ist in etwa gleichverteilt zwischen 90 und 140 Minuten.
- Sie können davon ausgehen, dass alle Besucher, die keine geführte Tour mitmachen, ein Headset haben wollen. Leider ist die Anzahl der Headsets beschränkt. Wir haben momentan nur 90 Headsets zur Verfügung.
- Falls alle Headsets entliehen sind, wartet der Besucher nicht darauf, dass wieder eines frei wird, sondern geht ohne Headset durch den Zoo. Ein Zoobesuch ohne Headset dauert etwa 60 Minuten.
- Nach der geführten Tour oder dem Besuch (mit oder ohne Headset), verlässt ungefähr die Hälfte der Besucher unmittelbar den Zoo. Die andere Hälfte geht noch in unseren Andenkenshop. Dort bleiben sie im Schnitt 15 Minuten und verlassen anschließend den Zoo. Den Bezahlvorgang im Shop müssen Sie nicht modellieren.

Skizzieren Sie ein entsprechendes ARENA-Modell zur Simulation der neuen Touristenattraktion. Benennen Sie alle Module aussagekräftig und notieren Sie zu treffende Einstellungen in den Flowchart- und Data-Modulen.

- **Flowchart Module**
  - Create: Zwischenankunftszeit | Anzahl Entitäten pro Ankunft
  - Dispose
  - Process: Seize-Delay-Release Typ | (Bearbeitungs-/ Delay-Zeit) | (Ressource(n))
  - Decide: by-condition + Bedingung/ by-chance + Wahrscheinlichkeiten
  - (Adjustable) Batch: Anzahl / (maximale Wartezeit)
  - Separate: Dublicate Original / Split Batch
  - Assign: Art | Name | Wert
  - Record: Count / Time Interval
- **Data Module**
  - Ressource: Ressource | Anzahl

Bitte beachten Sie, dass Sie **nicht** alle der hier aufgeführten Flowchart-Module in Ihrem Modell verwenden müssen und hier nicht aufgeführte Module nicht beschrieben werden müssen. Sie müssen desweiteren keine Angaben zum Run-Setup machen und können davon ausgehen, dass als Standard-Zeiteinheit Minuten eingestellt ist. Dokumentieren Sie (falls nötig) getroffene Annahmen.

b) Erläutern Sie kurz den Unterschied zwischen den Attributen und Variablen in ARENA.