Klausur Statistik (7.5 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 13.02.2019
Matrikelnummer			14:00 - 16:00 Uhr
Studienrichtung		**Sitzplatznummer**	
Semesterzahl		**Raum**	
Email (optional)			

Hinweis: Aufgabenblätter nicht auseinandertrennen!

Ergebnis:

Statistik	
Aufgabe	**Punkte**
1	
2	
3	
4	
Summe	
Note:	

Unterschrift des Kandidaten:____________________

Unterschrift des Prüfers:____________________

Hilfsmittel: Es gelten folgende Regelungen zu den erlaubten Hilfsmitteln:

- Nicht programmierbarer Taschenrechner
- Die vom Lehrstuhl offiziell herausgegebene Formelsammlung, 2. Auflage, (DIN A5, gebunden, orangener Umschlag), es sind keine weiteren Eintragungen oder Markierungen darin erlaubt. Ausgenommen sind farbliche Hinterlegungen von Textpassagen und/oder Formeln.
- R Reference Card von Jonathan Baron, es sind keine weiteren Eintragungen oder Markierungen darin erlaubt.

Bewertung: Für jede Aufgabe werden maximal zehn Punkte vergeben. Bewertet werden grundsätzlich nur Lösungen, die im Lösungsteil stehen und für die Folgendes beachtet wird:

- Der Lösungsweg muss aus einer Darstellung der einzelnen Rechenschritte ersichtlich sein.
- Antworten sind stets zu begründen, es sei denn es wird ausdrücklich keine Begründung verlangt.
- Unleserliche Aufgabenteile werden mit 0 Punkten bewertet.

Viel Erfolg!

Aufgabe 1 von 4

Die Seehundstation *Seal Sanctuary* an der Westküste Californiens hat sich der Aufzucht und Pflege junger Seehunde verschrieben. Um Aufschlüsse über den Zustand neu aufgefundener Jungtiere zu erhalten, werden deren Gewicht und Größe untersucht. Aus langjähriger Erfahrung wissen die Mitarbeiter, dass das Gewicht Y (in kg) von jungen Seehunden unabhängig und identisch normalverteilt ist mit Mittelwert 10 kg und Varianz 2 kg^2.

1. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig gefundener junger Seehund weniger als 8000 g wiegt?

2. Wie groß ist die Wahrscheinlichkeit, dass ein junger Seehund zwischen 8 und 12 kg wiegt?

3. Wie groß ist die Wahrscheinlichkeit, dass zwei junge Seehunde ein unterschiedliches Gewicht aufweisen?

4. Wie groß ist der Erwartungswert der Zufallsvariable $(Y - 10)/\sqrt{2}$?

5. Bestimmen Sie das 60%-Quantil von Y.

6. Bestimmen Sie, unter Berücksichtigung der Normalverteilungsannahme, den minimalen Umfang n für eine Zufallsstichprobe $Y_1, \dots, Y_n$ aus den Gewichten der jungen Seehunde. Hierbei soll die Wahrscheinlichkeit, dass das Stichprobenmittel $\bar{Y}_n$ um höchstens 0.05 vom wahren Erwartungswert $E(Y)$ abweicht, 95% betragen.

Ein Mitarbeiter der Seehundstation findet am Strand zwei Seehunde. Um die Entwicklung der beiden Tiere mit unabhängig und identisch verteilten Gewichten, beschrieben durch die Zufallsvariablen Y_1 und Y_2, zu vergleichen, beobachtet der Mitarbeiter die Zufallsvariable $Y_Z = Y_1 - Y_2$.

7. Geben Sie die Verteilungsfamilie, sowie Erwartungswert und Varianz von Y_Z an.

Gehen Sie nun davon aus, dass die beiden Zufallsvariablen Y_1 und Y_2 eine Kovarianz von $\mathrm{Cov}(Y_1, Y_2) = 1.5$ besitzen.

8. Berechnen Sie die Varianz von Y_Z.

9. Berechnen Sie den Korrelationskoeffizienten zwischen Y_1 und Y_2.

Aktuell befinden sich 16 Seehunde in der Station. Im Folgenden sind die absoluten Häufigkeiten für deren Alter tabelliert.

Alter des Seehundes	1	2	3	4	5	6	7	8	9	10
Absolute Häufigkeit	5	3	2	1	3	0	0	1	0	1

10. Bestimmen Sie den Modus des Merkmals “Alter des Seehundes”.

11. Geben Sie für das Merkmal “Alter des Seehundes” den Wert $x_{(0.5)}$ mit $x_{(0.5)} = \min\{x | F_n(x) \geq 0.5\}$ an.

12. Berechnen Sie den arithmetischen Mittelwert des Merkmals “Alter des Seehundes”.

13. Sei G: “Einschätzung des Gesundheitszustands” eines Seehundes. Der Gesundheitszustand wurde bewertet in den Stufen “schlecht”, “mittel”, “gut”, und “sehr gut”. Geben Sie das maximale Skalenniveau von G an.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

Sie denken darüber nach, sich einen gebrauchten Roller zu kaufen. Um die Folgekosten für Ersatzteile abschätzen zu können, interessieren Sie sich zunächst für die Zufallsvariable

$$Y: \text{“Anzahl der Teile, die im Laufe eines Jahres ausfallen”.}$$

Sie gehen davon aus, dass Y poissonverteilt ist mit Parameter $\lambda = 3.5$ (kurz: $Y \sim Pois(3.5)$).

1. Geben Sie die Wahrscheinlichkeit $P(Y = 4)$ an.

2. Geben Sie die Wahrscheinlichkeit $P(Y \geq 4)$ an.

3. Geben Sie den Modus der Verteilung an.

4. Geben Sie alle Zahlen aus der Menge $\{0;\ -1;\ 1.5;\ 10;\ -4;\ 12\}$ an, die im Träger von Y enthalten sind.

Ersatzteile gibt es in der Preiskategorie I für 3 Euro pro Stück und in der Preiskategorie II für 13 Euro pro Stück. Gegeben sind die folgenden zwei Zufallsvariablen:

B : "Teil aus Preiskategorie I fällt innerhalb eines Jahres aus" mit $B \sim Pois(0.4)$

T : "Teil aus Preiskategorie II fällt innerhalb eines Jahres aus" mit $T \sim Pois(0.2)$.

Insgesamt hat der Roller 40 Teile aus Preiskategorie I und 30 Teile aus Preiskategorie II verbaut, wobei die Teile jeweils stochastisch unabhängig voneinander ausfallen. Die Zufallsvariable K: "Gesamtkosten für Ersatzteile pro Jahr" ist somit gegeben durch:

$$K = 3\sum_{i=1}^{40} B_i + 13\sum_{i=1}^{30} T_i.$$

5. Berechnen Sie den Erwartungswert von K.

6. Berechnen Sie die Varianz von K.

Außerdem betrachten Sie die Zufallsvariable

X : "Monate bis zum Defekt des Motors".

Gehen Sie davon aus, dass X Weibull-verteilt ist mit zunächst unbekannten Parametern λ und r. Für $x \geq 0$ ist die Dichte somit gegeben durch

$$f(x;\lambda,r) = r\lambda(\lambda x)^{r-1}\exp(-(\lambda x)^r)$$

und die Verteilungsfunktion durch

$$F(x;\lambda,r) = 1 - \exp(-(\lambda x)^r).$$

Nehmen Sie im Folgenden an, dass $r = 1$ gilt.

7. Zeigen Sie, dass die Log-Likelihood Funktion für Beobachtungen $x_1, ..., x_n$ die folgende Form hat:
$$LL(x_1, ..., x_n; \lambda) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

8. Zeigen Sie, dass der Maximum-Likelihood Schätzer für λ durch $\hat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$ gegeben ist. (Hinweis: Gehen Sie davon aus, dass die hinreichende Bedingung für ein Maximum erfüllt ist.)

9. Für die Zufallsvariable X ziehen Sie die folgende einfache Stichprobe:

i	1	2	3	4	5	6
Monate x_i	30.4	20.2	77.1	45.2	69.1	42.3

Berechnen Sie den Maximum-Likelihood Schätzer $\hat{\lambda}_{ML}$.

Nehmen Sie nun an, dass $\lambda = 2$ und $r = 1$ gilt, d.h.

$$f(x) = 2\exp(-2x)$$

und

$$F(x) = 1 - \exp(-2x).$$

10. Berechnen Sie die Hazardrate.

11. Geben Sie den Wert der Dichte an der Stelle $x = 1$ an.

12. Sie haben verschiedene Zahlen in die Verteilungsfunktion eingesetzt und folgende Ergebnisse erhalten:

$$F(0) = 0,\ F(0.25) = 0.3935,\ F(0.5) = 0.6321,\ F(1) = 0.8647,\ F(2) = 0.9817.$$

Das 80% Quantil der Verteilung liegt somit in welchem der vier betrachteten Intervalle: [0;0.25], [0.25;0.5], [0.5;1] oder [1;2]?

13. Wie bezeichnet man eine Weibull-Verteilung mit Parameter $r = 1$?

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

In der folgenden Tabelle sind die 12-Uhr Temperaturwerte (in $°C$) der ersten Dezemberwoche in einer Stadt dargestellt.

Beobachtung i	1	2	3	4	5	6	7
Datum	01.12.	02.12.	03.12.	04.12.	05.12.	06.12.	07.12.
Messwert x_i	3.5	2.3	-0.4	2.1	1.3	-1.1	-7.9

1. Berechnen Sie die Durchschnittstemperatur der ersten Dezemberwoche als arithmetisches Mittel.

In der folgenden Tabelle ist die jährliche Veränderung des Bruttoinlandsprodukts eines Landes dargestellt.

Beobachtung i	1	2	3	4	5
Jahr	2014	2015	2016	2017	2018
Wachstumsrate r_i	0.03	0.01	-0.01	0.02	0.01

2. Betrachten Sie die Wachstumsfaktoren $1 + r_i$ und berechnen Sie den durchschnittlichen Wachstumsfaktor des Bruttoinlandsprodukts des Landes als geometrisches Mittel.

Ein Zug fährt auf drei Teilstrecken unterschiedlicher Länge s_i mit unterschiedlichen Durchschnittsgeschwindigkeiten v_i.

Teilstrecke i	1	2	3
Streckenlänge s_i	15 km	40 km	20 km
Durchschnittsgeschwindigkeit v_i	50 km/h	160 km/h	70 km/h

3. Berechnen Sie die Durchschnittsgeschwindigkeit für die gesamte Strecke als harmonisches Mittel.

Die folgende Tabelle enthält statistische Kennzahlen von Stichproben zur Verteilung der Körpergröße in zwei Ländern A und B. Nehmen Sie an, dass es sich dabei um Realisationen normalverteilter Grundgesamtheiten mit unbekannten, homogenen Varianzen handelt.

Land	Stichprobenmittel	Stichprobenvarianz	Stichprobengröße
Land A	175.0	55.2	101
Land B	169.0	34.1	81

4. Geben Sie die Realisation von Ober- und Untergrenze des zweiseitig symmetrischen 95% Konfidenzintervalls für den Mittelwert der Stichprobe von Land A an.

5. Nehmen Sie nun an, dass das realisierte, symmetrische 95% Konfidenzintervall für den Mittelwert der Stichprobe von Land B durch $[167; 171]$ gegeben ist. Begründen Sie kurz, ob die zugehörige Nullhypothese $H_0 : \mu_B = 172$ auf einem Signifikanzniveau von 5% abgelehnt werden kann oder nicht.

Sie interessieren sich nun für die Mittelwertdifferenz mit der Hypothese $H_0 : \mu_A - \mu_B \leq 0$.

6. Nehmen Sie an, dass die gepoolte Stichprobenvarianz $\overline{S}^2 = 122.3$ beträgt. Berechnen Sie die Realisation der zugehörigen Teststatistik und geben Sie deren theoretische Verteilung unter der Nullhypothese inklusive Parameter an.

7. Nehmen Sie nun an, dass der p-Wert des Tests 0.3012 beträgt. Begründen Sie kurz, ob die Nullhypothese für ein Signifikanzniveau von 5% abgelehnt werden kann oder nicht.

Sie betrachten nun zwei normalverteilte Zufallsvariablen $X \sim \mathcal{N}(0,1)$ und $Y \sim \mathcal{N}(0,1)$.

8. Wie ist die Zufallsvariable X^2 verteilt? Geben Sie sowohl Verteilungsfamilie, als auch Parameter konkret an.

9. Wie ist die Zufallsvariable $X^2 + Y^2$ verteilt? Geben Sie sowohl Verteilungsfamilie, als auch Parameter konkret an.

10. Sie möchten eine binomialverteilte Zufallsvariable B mit Parametern $n = 10$ und $p = 0.3$ durch eine normalverteilte Zufallsvariable approximieren $B \overset{approx.}{\sim} \mathcal{N}(\mu, \sigma^2)$ und setzen dazu die ersten beiden Momente der Verteilungen gleich. Wie lautet der Wert für den Parameter μ konkret?

Die Zufallsvariable Z besitzt die Verteilungsfunktion $F(z) = e^{-e^{-z}}$.

11. Geben Sie die Quantilsfunktion $F^{-1}(u)$ von Z als Umkehrfunktion der Verteilungsfunktion $F(z)$ an. Ersetzen Sie dazu $F(z)$ durch u und z durch $F^{-1}(u)$ und lösen Sie den Ausdruck anschließend nach $F^{-1}(u)$ auf.

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

Sie besitzen einen Elektromarkt und verkaufen Kaffeemaschinen verschiedener Hersteller. Sie erhalten über Ihre Service-Hotline verschiedene Anruf-Arten. Sei die Zufallsvariable B: 'Anzahl der Beschwerde-Anrufe von n Anrufen' binomialverteilt, wobei $p = 0.35$.

1. Wie groß ist die Wahrscheinlichkeit, dass Sie von den nächsten 15 Anrufen genau 13 Beschwerden entgegennehmen?

2. Wie groß ist die Wahrscheinlichkeit, dass Sie von den nächsten 15 Anrufen mindestens eine und weniger als 10 Beschwerden entgegennehmen?

3. Wieviele Beschwerdeanrufe erwarten Sie bei insgesamt 5000 entgegengenommenen Anrufen?

Sie verkaufen Kaffeemaschinen der drei Hersteller A, E und I. Sie prüfen die Geräte bei Anlieferung auf die drei Zustände 'defekt', 'kleinere Mängel' und 'guter Zustand'.
Sie vermuten, dass die Variablen Z: 'Zustand eines Gerätes' und H: 'Hersteller eines Gerätes' nicht unabhängig voneinander sind. Dies wollen Sie mit einer Irrtumswahrscheinlichkeit von $\alpha = 10\%$ testen.

4. Beschreiben Sie verbal die beiden Fehler, die bei Testentscheidungen unter Unsicherheit auftreten können.

Sie führen einen χ^2-Unabhängigkeitstest durch.

5. Geben Sie die theoretische Teststatistik inklusive asymptotischer Verteilung unter der Nullhypothese an.

6. Wie lautet die kritische Schranke mit konkretem Wert?

Sei der Wert der realisierten Teststatistik $\chi^2 = 2$.
Verwenden Sie im folgenden als Wert der kritischen Schranke $KS = 10$.

7. Treffen Sie eine Testentscheidung und begründen Sie diese.

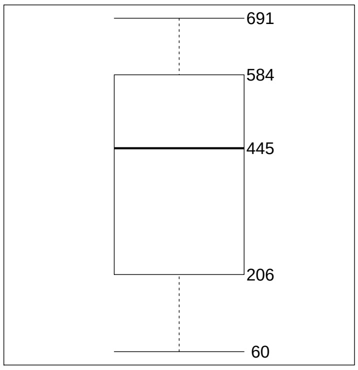
8. Würden Sie für den von Ihnen durchgeführten Test einen p-Wert von 0.008, 0.08 oder 0.8 erwarten? Begründen Sie Ihre Antwort.

9. Welchen Wert nimmt die mittlere quadratische Kontingenz in diesem Kontext bei vollständiger Abhängigkeit an?

Sei die mittlere quadratische Kontingenz $\phi^2 = 0.0033$.

10. Berechnen Sie Cramér's V und interpretieren Sie Ihr Ergebnis im Kontext.

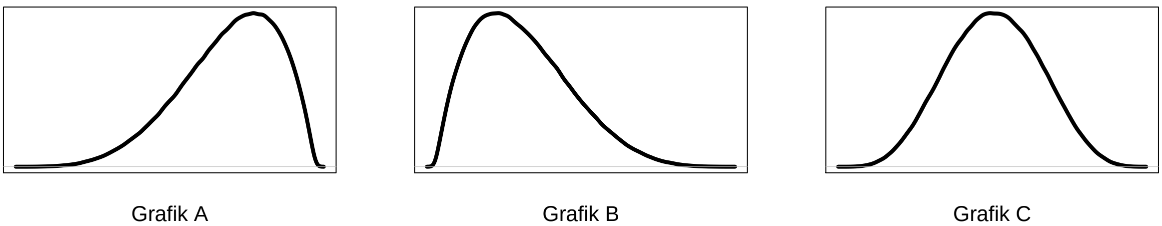
Sie beschäftigen sich nun mit zurückgesendeter Ware. Dazu betrachten Sie die Zufallsvariable T: 'Tage bis zur Rücksendung einer Ware' mit $\mu_T = 388$ und $\sigma_T = 154$. Außerdem haben Sie folgenden Box-Plot vorliegen:



11. Geben Sie den Quartilsabstand an.

12. Berechnen Sie den zweiten Pearsonschen Schiefekoeffizienten.

13. Welche der folgenden Grafiken skizziert am ehesten die Verteilung von T?



Schmierpapier zu Aufgabe 4

Klausur Statistik (10 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 13.02.2019
Matrikelnummer			14:00 - 16:00 Uhr
Studienrichtung		**Sitzplatznummer**	
Semesterzahl		**Raum**	
Email (optional)			

Hinweis: Aufgabenblätter nicht auseinandertrennen!

Ergebnis:

Statistik	
Aufgabe	**Punkte**
1	
2	
3	
4	
Summe	
Note:	

Unterschrift des Kandidaten:____________________

Unterschrift des Prüfers:____________________

Hilfsmittel: Es gelten folgende Regelungen zu den erlaubten Hilfsmitteln:

- Nicht programmierbarer Taschenrechner
- Die vom Lehrstuhl offiziell herausgegebene Formelsammlung, 2. Auflage, (DIN A5, gebunden, orangener Umschlag), es sind keine weiteren Eintragungen oder Markierungen darin erlaubt. Ausgenommen sind farbliche Hinterlegungen von Textpassagen und/oder Formeln.
- R Reference Card von Jonathan Baron, es sind keine weiteren Eintragungen oder Markierungen darin erlaubt.

Bewertung: Für jede Aufgabe werden maximal zehn Punkte vergeben. Bewertet werden grundsätzlich nur Lösungen, die im Lösungsteil stehen und für die Folgendes beachtet wird:

- Der Lösungsweg muss aus einer Darstellung der einzelnen Rechenschritte ersichtlich sein.
- Antworten sind stets zu begründen, es sei denn es wird ausdrücklich keine Begründung verlangt.
- Unleserliche Aufgabenteile werden mit 0 Punkten bewertet.

Viel Erfolg!

Aufgabe 1 von 4

Die Seehundstation *Seal Sanctuary* an der Westküste Californiens hat sich der Aufzucht und Pflege junger Seehunde verschrieben. Um Aufschlüsse über den Zustand neu aufgefundener Jungtiere zu erhalten, werden deren Gewicht und Größe untersucht. Aus langjähriger Erfahrung wissen die Mitarbeiter, dass das Gewicht Y (in kg) von jungen Seehunden unabhängig und identisch normalverteilt ist mit Mittelwert 10 kg und Varianz 2 kg^2.

1. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig gefundener junger Seehund weniger als 8000 g wiegt?

2. Wie groß ist die Wahrscheinlichkeit, dass ein junger Seehund zwischen 8 und 12 kg wiegt?

3. Wie groß ist die Wahrscheinlichkeit, dass zwei junge Seehunde ein unterschiedliches Gewicht aufweisen?

4. Wie groß ist der Erwartungswert der Zufallsvariable $(Y - 10)/\sqrt{2}$?

5. Bestimmen Sie das 60%-Quantil von Y.

6. Bestimmen Sie, unter Berücksichtigung der Normalverteilungsannahme, den minimalen Umfang n für eine Zufallsstichprobe $Y_1, \ldots, Y_n$ aus den Gewichten der jungen Seehunde. Hierbei soll die Wahrscheinlichkeit, dass das Stichprobenmittel $\bar{Y}_n$ um höchstens 0.05 vom wahren Erwartungswert $E(Y)$ abweicht, 95% betragen.

Ein Mitarbeiter der Seehundstation findet am Strand zwei Seehunde. Um die Entwicklung der beiden Tiere mit unabhängig und identisch verteilten Gewichten, beschrieben durch die Zufallsvariablen Y_1 und Y_2, zu vergleichen, beobachtet der Mitarbeiter die Zufallsvariable $Y_Z = Y_1 - Y_2$.

7. Geben Sie die Verteilungsfamilie, sowie Erwartungswert und Varianz von Y_Z an.

Gehen Sie nun davon aus, dass die beiden Zufallsvariablen Y_1 und Y_2 eine Kovarianz von $\text{Cov}(Y_1, Y_2) = 1.5$ besitzen.

8. Berechnen Sie die Varianz von Y_Z.

9. Berechnen Sie den Korrelationskoeffizienten zwischen Y_1 und Y_2.

Aktuell befinden sich 16 Seehunde in der Station. Im Folgenden sind die absoluten Häufigkeiten für deren Alter tabelliert.

Alter des Seehundes	1	2	3	4	5	6	7	8	9	10
Absolute Häufigkeit	5	3	2	1	3	0	0	1	0	1

10. Bestimmen Sie den Modus des Merkmals “Alter des Seehundes”.

11. Geben Sie für das Merkmal “Alter des Seehundes” den Wert $x_{(0.5)}$ mit $x_{(0.5)} = \min\{x | F_n(x) \geq 0.5\}$ an.

12. Berechnen Sie den arithmetischen Mittelwert des Merkmals “Alter des Seehundes”.

13. Sei G: “Einschätzung des Gesundheitszustands” eines Seehundes. Der Gesundheitszustand wurde bewertet in den Stufen “schlecht”, “mittel”, “gut”, und “sehr gut”. Geben Sie das maximale Skalenniveau von G an.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

Sie denken darüber nach, sich einen gebrauchten Roller zu kaufen. Um die Folgekosten für Ersatzteile abschätzen zu können, interessieren Sie sich zunächst für die Zufallsvariable

$$Y : \text{“Anzahl der Teile, die im Laufe eines Jahres ausfallen”}.$$

Sie gehen davon aus, dass Y poissonverteilt ist mit Parameter $\lambda = 3.5$ (kurz: $Y \sim Pois(3.5)$).

1. Geben Sie die Wahrscheinlichkeit $P(Y = 4)$ an.

2. Geben Sie die Wahrscheinlichkeit $P(Y \geq 4)$ an.

3. Geben Sie den Modus der Verteilung an.

4. Geben Sie alle Zahlen aus der Menge $\{0;\ -1;\ 1.5;\ 10;\ -4;\ 12\}$ an, die im Träger von Y enthalten sind.

Ersatzteile gibt es in der Preiskategorie I für 3 Euro pro Stück und in der Preiskategorie II für 13 Euro pro Stück. Gegeben sind die folgenden zwei Zufallsvariablen:

B : "Teil aus Preiskategorie I fällt innerhalb eines Jahres aus" mit $B \sim Pois(0.4)$

T : "Teil aus Preiskategorie II fällt innerhalb eines Jahres aus" mit $T \sim Pois(0.2)$.

Insgesamt hat der Roller 40 Teile aus Preiskategorie I und 30 Teile aus Preiskategorie II verbaut, wobei die Teile jeweils stochastisch unabhängig voneinander ausfallen. Die Zufallsvariable K: "Gesamtkosten für Ersatzteile pro Jahr" ist somit gegeben durch:

$$K = 3\sum_{i=1}^{40} B_i + 13\sum_{i=1}^{30} T_i.$$

5. Berechnen Sie den Erwartungswert von K.

6. Berechnen Sie die Varianz von K.

Außerdem betrachten Sie die Zufallsvariable

X : "Monate bis zum Defekt des Motors".

Gehen Sie davon aus, dass X Weibull-verteilt ist mit zunächst unbekannten Parametern λ und r. Für $x \geq 0$ ist die Dichte somit gegeben durch

$$f(x; \lambda, r) = r\lambda(\lambda x)^{r-1} \exp(-(\lambda x)^r)$$

und die Verteilungsfunktion durch

$$F(x; \lambda, r) = 1 - \exp(-(\lambda x)^r).$$

Nehmen Sie im Folgenden an, dass $r = 1$ gilt.

7. Zeigen Sie, dass die Log-Likelihood Funktion für Beobachtungen $x_1, ..., x_n$ die folgende Form hat:

$$LL(x_1, ..., x_n; \lambda) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

8. Zeigen Sie, dass der Maximum-Likelihood Schätzer für λ durch $\hat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$ gegeben ist. (Hinweis: Gehen Sie davon aus, dass die hinreichende Bedingung für ein Maximum erfüllt ist.)

9. Für die Zufallsvariable X ziehen Sie die folgende einfache Stichprobe:

i	1	2	3	4	5	6
Monate x_i	30.4	20.2	77.1	45.2	69.1	42.3

Berechnen Sie den Maximum-Likelihood Schätzer $\hat{\lambda}_{ML}$.

Nehmen Sie nun an, dass $\lambda = 2$ und $r = 1$ gilt, d.h.

$$f(x) = 2\exp(-2x)$$

und

$$F(x) = 1 - \exp(-2x).$$

10. Berechnen Sie die Hazardrate.

11. Geben Sie den Wert der Dichte an der Stelle $x = 1$ an.

12. Sie haben verschiedene Zahlen in die Verteilungsfunktion eingesetzt und folgende Ergebnisse erhalten:

    $$F(0) = 0,\ F(0.25) = 0.3935,\ F(0.5) = 0.6321,\ F(1) = 0.8647,\ F(2) = 0.9817.$$

    Das 80% Quantil der Verteilung liegt somit in welchem der vier betrachteten Intervalle: [0;0.25], [0.25;0.5], [0.5;1] oder [1;2]?

13. Wie bezeichnet man eine Weibull-Verteilung mit Parameter $r = 1$?

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

In der folgenden Tabelle sind die 12-Uhr Temperaturwerte (in °C) der ersten Dezemberwoche in einer Stadt dargestellt.

Beobachtung i	1	2	3	4	5	6	7
Datum	01.12.	02.12.	03.12.	04.12.	05.12.	06.12.	07.12.
Messwert x_i	3.5	2.3	-0.4	2.1	1.3	-1.1	-7.9

1. Berechnen Sie die Durchschnittstemperatur der ersten Dezemberwoche als arithmetisches Mittel.

In der folgenden Tabelle ist die jährliche Veränderung des Bruttoinlandsprodukts eines Landes dargestellt.

Beobachtung i	1	2	3	4	5
Jahr	2014	2015	2016	2017	2018
Wachstumsrate r_i	0.03	0.01	-0.01	0.02	0.01

2. Betrachten Sie die Wachstumsfaktoren $1 + r_i$ und berechnen Sie den durchschnittlichen Wachstumsfaktor des Bruttoinlandsprodukts des Landes als geometrisches Mittel.

Ein Zug fährt auf drei Teilstrecken unterschiedlicher Länge s_i mit unterschiedlichen Durchschnittsgeschwindigkeiten v_i.

Teilstrecke i	1	2	3
Streckenlänge s_i	15 km	40 km	20 km
Durchschnittsgeschwindigkeit v_i	50 km/h	160 km/h	70 km/h

3. Berechnen Sie die Durchschnittsgeschwindigkeit für die gesamte Strecke als harmonisches Mittel.

Die folgende Tabelle enthält statistische Kennzahlen von Stichproben zur Verteilung der Körpergröße in zwei Ländern A und B. Nehmen Sie an, dass es sich dabei um Realisationen normalverteilter Grundgesamtheiten mit unbekannten, homogenen Varianzen handelt.

Land	Stichprobenmittel	Stichprobenvarianz	Stichprobengröße
Land A	175.0	55.2	101
Land B	169.0	34.1	81

4. Geben Sie die Realisation von Ober- und Untergrenze des zweiseitig symmetrischen 95% Konfidenzintervalls für den Mittelwert der Stichprobe von Land A an.

5. Nehmen Sie nun an, dass das realisierte, symmetrische 95% Konfidenzintervall für den Mittelwert der Stichprobe von Land B durch $[167; 171]$ gegeben ist. Begründen Sie kurz, ob die zugehörige Nullhypothese $H_0 : \mu_B = 172$ auf einem Signifikanzniveau von 5% abgelehnt werden kann oder nicht.

Sie interessieren sich nun für die Mittelwertdifferenz mit der Hypothese $H_0 : \mu_A - \mu_B \leq 0$.

6. Nehmen Sie an, dass die gepoolte Stichprobenvarianz $\overline{S}^2 = 122.3$ beträgt. Berechnen Sie die Realisation der zugehörigen Teststatistik und geben Sie deren theoretische Verteilung unter der Nullhypothese inklusive Parameter an.

7. Nehmen Sie nun an, dass der p-Wert des Tests 0.3012 beträgt. Begründen Sie kurz, ob die Nullhypothese für ein Signifikanzniveau von 5% abgelehnt werden kann oder nicht.

Sie betrachten nun zwei normalverteilte Zufallsvariablen $X \sim \mathcal{N}(0,1)$ und $Y \sim \mathcal{N}(0,1)$.

8. Wie ist die Zufallsvariable X^2 verteilt? Geben Sie sowohl Verteilungsfamilie, als auch Parameter konkret an.

9. Wie ist die Zufallsvariable $X^2 + Y^2$ verteilt? Geben Sie sowohl Verteilungsfamilie, als auch Parameter konkret an.

10. Sie möchten eine binomialverteilte Zufallsvariable B mit Parametern $n = 10$ und $p = 0.3$ durch eine normalverteilte Zufallsvariable approximieren $B \overset{approx.}{\sim} \mathcal{N}(\mu, \sigma^2)$ und setzen dazu die ersten beiden Momente der Verteilungen gleich. Wie lautet der Wert für den Parameter μ konkret?

Die Zufallsvariable Z besitzt die Verteilungsfunktion $F(z) = e^{-e^{-z}}$.

11. Geben Sie die Quantilsfunktion $F^{-1}(u)$ von Z als Umkehrfunktion der Verteilungsfunktion $F(z)$ an. Ersetzen Sie dazu $F(z)$ durch u und z durch $F^{-1}(u)$ und lösen Sie den Ausdruck anschließend nach $F^{-1}(u)$ auf.

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

In Ihrem Workspace liegt ein bereits erstellter Dataframe `dax`, welcher den täglichen Kursverlauf des Deutschen Aktienindex (DAX) im Zeitraum von 2000 bis 2018 beinhaltet.[1] Der Dataframe enthält folgende Merkmale:

Ausprägung Merkmal	**Spaltenname**
Schlusskurs:	`Close`
Eröffnungskurs:	`Open`
Niedrigster Kurs:	`Low`
Höchster Kurs:	`High`
Handelsvolumen:	`Volume`

Arbeiten Sie zunächst mit folgendem `R`-Output:

```
'dax':        4703 obs. of  5 variables:
 $ Close :  num  6750.76 6586.95 6502.07 6474.92 6780.96 ...
 $ Open  :  num  6961.72 6747.24 6585.85 6501.45 6489.94 ...
 $ Low   :  num  6720.87 6510.46 6388.91 6402.63 6470.14 ...
 $ High  :  num  7159.33 6755.36 6585.85 6539.31 6791.53 ...
 $ Volume:  num  43072500 46678400 52682800 41180600 56058900 ...
```

Hinweis: Spezifizieren Sie die Befehle so, dass eine Installation zusätzlicher Pakete nicht notwendig ist.

1. Geben Sie den `R`-Befehl an, mit dem Sie die Spannweite des Handelsvolumens erhalten.

2. Geben Sie den `R`-Befehl an, mit dem Sie die Anzahl der Tage mit einem Schlusskurs über 6000 erhalten.

[1] https://de.finance.yahoo.com

3. Welche Outputwerte liefert der Befehl `dax$Low[seq(2,4,2)]`?

Folgender R-Code wird nun in die Konsole eingegeben:

```
> a <- mean(dax$Open^2)-mean(dax$Open)^2
> b <- 1/(a^(-0.5))
```

4. Welche statistischen Kennzahlen beschreiben die Variablen a und b in obigem Befehl?

5. Geben Sie den R-Befehl für das Erstellen eines logischen Vektors `niedriger` an, der für jedes Element den Wert `TRUE` enthält, falls der tägliche Schlusskurs unter dem Eröffnungskurs liegt.

6. Mit welchem R-Code können Sie sich den Teildatensatz anzeigen lassen, der lediglich die Daten enthält, bei denen der Höchstkurs niedriger als 5000 liegt?

7. Nehmen Sie zu folgender Aussage Stellung: "Die betragsmäßige prozentuale Veränderung des Schlusskurses zwischen dem ersten und zweiten Tag ist größer als die betragsmäßige prozentuale Veränderung zwischen dem dritten und vierten Tag". **Mit Begründung!**

Aufgrund von Risikodiversifizierung wird im Folgenden ein Portfolio P aus Aktien betrachtet. Dieses ist definiert als:

$$P = aX + (1-a)Y, \quad a \in [0,1].$$

Die Zufallsvariable X beschreibt die Rendite von Aktie 1, wobei X den Erwartungswert $e1$ und Varianz $var1$ besitzt. Die Rendite von Aktie 2 wird durch die Zufallsvariable Y mit Erwartungswert $e2$ und Varianz $var2$ beschrieben. Die Kovarianz zwischen den Zufallsvariablen X und Y ist mit cov gekenntzeichnet. a beschreibt den relativen Anteil von Aktie 1 im Portfolio P.

8. Geben Sie den `R`-Befehl an, mit dem Sie eine Funktion `e` implementieren, welche als Inputargumente $a, e1, e2$ besitzt und den Erwartungswert von P zurückgibt.

Folgender R-Code wird nun in die Konsole eingegeben:

```
> j = 0
> i = 0
> while(i < 3)
        {j = j + i
        i = i+1}
```

9. Welchen Wert erhalten Sie für j?

10. Vervollständigen Sie folgenden R-Code, um die logarithmierte Likelihoodfunktion der Exponentialverteilung mit Parameter `lambda` in R zu berechnen.

```
LL_exp = function(       ,data){
    sum(         (          (data,        )         )         )
}
```

Schmierpapier zu Aufgabe 4

Lösung WS 2018/2019 Aufgabe 1

1. $P(Y < 8) = 1 - \Phi(1.41) = 1 - 0.9207 = 0.0793$ 1 P
2. $P(8 < Y < 12) = 2\Phi(1.41) - 1 = 2 \cdot 0.9207 - 1 = 0.8414$ 1 P
3. 1 0.5 P
4. 0 0.5 P
5. $x_{(0.6)} = 0.2533 \cdot \sqrt{2} + 10 = 10.3582$ 1 P
6. $n \geq (\lambda_{1-\alpha/2}\frac{\sigma}{\epsilon})^2 = (1.96\frac{\sqrt{2}}{0.05})^2 = 3073.28 \rightarrow n \geq 3074$ 1 P
7. Normalverteilung mit $E[Y_Z] = 0$, $Var[Y_Z] = 4$ 1.5 P
8. $Var[Y_Z] = Var[Y_1] + Var[Y_2] + 2 \cdot (-1) \cdot 1.5 = 1.$ 1 P
9. $\rho_{Y_1Y_2} = \frac{1.5}{\sqrt{2 \cdot 2}} = 0.75.$ 0.5 P
10. 1 0.5 P
11. 2 0.5 P
12. $(5 \cdot 1 + ... + 1 \cdot 10)/16 = 3.375$ 0.5 P
13. Ordinalskala 0.5 P

Lösung WS 2018/2019 Aufgabe 2

1. 0.1888 0.5 P
2. $1 - P(X \leq 3) = 0.4634$ 1P

3.) 3 0.5P

4. 0,10,12 0.5P
5. $E[K] = 3\sum_{i=1}^{40} E[B_i] + 13\sum_{i=1}^{30} E[T_i] = 126$ 1P
6. $Var[K] = 3^2\sum_{i=1}^{40} Var(B_i) + 13^2\sum_{i=1}^{30} Var(T_i) = 1158$ 1P
7. $LL(\mathbf{x};\lambda) = \sum_{i=1}^{n} \log(\lambda\exp(-(\lambda x_i))) = \sum_{i=1}^{n} \log(\lambda) - \lambda x_i = n\log(\lambda) - \lambda\sum_{i=1}^{n} x_i$ 1.5 P
8. $\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$. 1 P
9. 0.0211 0.5P
10. $\frac{f(x)}{1-F(x)} = 2$ 1P
11. 0.2707 0.5P
12. [0.5;1] 0.5P
13. Exponentialverteilt 0.5P

Lösung WS 2018/2019 Aufgabe 3

1. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = -0.0286[^\circ C]$ 1 P
2. $\bar{x}_G = \left(\prod_{i=1}^{n} 1 + r_i\right)^{1/n} = 1.01191$ 1 P
3. $\bar{x}_H = \frac{15km+40km+20km}{\frac{15km}{50km/h}+\frac{40km}{160km/h}+\frac{20km}{70km/h}} = 89.7436km/h$ 1 P
4. $[175 - t_{0.975,100}\frac{\sqrt{55.2}}{\sqrt{101}}, 175 + t_{0.975,100}\frac{\sqrt{55.2}}{\sqrt{101}}] = [173.5333, 176.4667]$ 1 P
5. H_0 kann verworfen werden, da $172 \notin [167, 171]$. 0.5 P
6. $t_{101,81} = \frac{175-169-0}{\sqrt{122.3}\sqrt{\frac{1}{101}+\frac{1}{81}}} = 3.6375, \quad T_{101,81} \sim t(101+81-2=180)$ 1.5 P
7. $p > \alpha \Rightarrow H_0$ nicht verwerfen. 0.5 P
8. $\mathcal{X}^2(1)$ 1 P
9. $\mathcal{X}^2(2)$ 1 P
10. $\mu = np = 3$ 0.5 P
11. $u = \exp(-\exp(-F^{-1}(u))) \Leftrightarrow F^{-1}(u) = -ln(-ln(u))$ 1 P

Lösung WS 2018/2019 Aufgabe 4 (10 ECTS)

1. `max(dax$Volume)-min(dax$Volume)` 1.0 P

2. `sum(dax$Close>6000)` 1.0 P

3. 6510.46 6402.63 1.0 P

4. Varianz, Standardabweichung 1.0 P

5. `niedriger<-dax$Close<dax$Open` 1.0 P

6. `dax[dax$High<5000,]` 1.0 P

7. Aussage richtig
   6586.95/6750.76-1= |-0.0243| = 0.0243, 6474.92/6502.07-1=|-0.0042| = 0.0042, 1.0 P

8. `e<-function(a,e1,e2){a*e1+(1-a)*e2}` 1.0 P

9. 3 1.0 P

10. `LL_exp=function(lambda,data){sum(log(dexp(data,lambda)))}` 1.0 P

Lösung WS 2018/2019 Aufgabe 4 (7.5 ECTS)

1. $f_{Bin}(13; n = 15; p = 0.35) = 0.0001$ 0.5 P

2. $F_{Bin}(9; n = 15; p = 0.35) - F_{Bin}(0; n = 15; p = 0.35)$
   $= 0.9876 - 0.0016 = 0.986$ 1 P

3. $5000 * 0.35 = 1750$ 0.5 P

4. Fehler 1. Art: H_0 ablehnen, obwohl H_0 richtig ist
   Fehler 2. Art: H_0 nicht ablehnen, obwohl H_0 falsch ist 1 P

5. Prüfgröße, Verteilung: $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(N_{ij} - \frac{N_{i.}N_{.j}}{n}\right)^2}{\frac{N_{i.}N_{.j}}{n}} \overset{a}{\sim} \chi^2((k-1)(l-1))$

   1 P

6. kritische Schranke: $\chi^2_{0.90}(4) = 7.78$ 0.5 P

7. Testentscheidung: Da $2 \not> 10$ 0.5 P
   Die Nullhypothese kann auf dem 10% Signifikanzniveau nicht abgelehnt werden.
   0.5 P

8. Erwarteter p-Wert: 0.8, da nicht auf dem 10% Signifikanzniveau abgelehnt werden konnte.
   1 P

9. ϕ^2 bei vollständiger Abhängigkeit: $min((k-1)(l-1)) = 2$ 0.5 P

10. Cramer's V: $\sqrt{\frac{\phi^2}{min((k-1)(l-1))}} = 0.0407$ 0.5 P
    Interpretation: Kaum Zusammenhang 0.5 P

11. Quartilsabstand: $584 - 206 = 378$ 0.5 P

12. Zweiter Pearsonscher Schiefekoeffizient:
    $\frac{\mu_T - \text{Median}}{\sigma_T} = \frac{388-445}{154} = -0.3701$ 1 P

13. Grafik A 0.5 P