Klausur Statistik (7,5 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 12.02.2014 14:00 – 16:00 Uhr
Matrikelnummer			
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: Zugelassen sind eine selbsterstellte maximal vierseitige (DIN A4 doppelseitig) Formelsammlung, die vom Lehrstuhl zum Download bereitgestellte Tabellensammlung ohne Eintragungen und die R-Reference-Card (by Jonathan Baron) ohne Eintragungen.
Darüber hinaus sind Taschenrechner zugelassen. Es dürfen jedoch keine Programme oder Programmteile verwendet werden, die nicht fest in den Taschenrechner eingebaut sind. Alle Hilfsmittel sind selbst mitzubringen.

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

In einer Umfrage zum Rauchverhalten wurden insgesamt 1000 Personen interviewt, die in Deutschland leben und mindestens 18 Jahre alt sind. Folgende Statistik zeigt den Anteil von Rauchern, Ex-Rauchern und Nie-Rauchern (Personen, die nie geraucht haben) aufgeteilt nach Geschlecht. Bekannt ist noch, dass $f(R|F) = 26.2\%$ ist.

Rauchverhalten → Geschlecht ↓	Raucher (R)	Ex-Raucher (E)	Nie-Raucher (N)	
Männer (M)				43.4%
Frauen (F)				
	30%	26.6%	43.4%	1

1. Berechnen Sie den Anteil der Raucherinnen. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 14.8%.)

2. Berechnen Sie $f(N \cap M)$, wenn Sie wissen, dass $f(E \cap M) = f(N \cap M)$ ist.

3. Wie viele der Befragten haben noch nie geraucht?

4. Interpretieren Sie $n(F \cap N) = 293$.

5. Um die Stärke der Abhängigkeit zwischen „Rauchverhalten” und „Geschlecht” zu untersuchen, wurde die mittlere quadratische Kontingenz berechnet. Diese beträgt 0.0379. Berechnen und interpretieren Sie das Cramérsche Kontingenzmaß V für die beiden Merkmale.

6. Nennen Sie ein geeignetes Streuungsmaß für das Merkmal „Rauchverhalten”.

7. Nennen Sie ein geeignetes Lagemaß für das Merkmal „Geschlecht” und bestimmen Sie dieses.

8. Von den 1000 Untersuchungsteilnehmern werden nun 20 Personen zufällig ausgewählt und befragt (Doppelbefragungen seien zugelassen). Wie groß ist die Wahrscheinlichkeit, dass genau die Hälfte der 20 Befragten Raucher/innen sind, wenn der Prozentsatz der Raucher in der Grundgesamtheit genauso groß ist wie der Anteil der Raucher an den Untersuchungsteilnehmern?

9. Um welche Verteilung handelt es sich, wenn eine Person von den 1000 Personen befragt wird, ob er/sie raucht?

Folgende Daten geben den durchschnittlichen täglichen Zigarettenkonsum von 10 Personen an.

Person i	1	2	3	4	5	6	7	8	9	10
Anzahl der gerauchten Zigaretten pro Tag (Z)	1	3	4	8	8	15	21	21	10	7

10. Bestimmen Sie das untere Quartil der empirischen Verteilung von Z.

11. Berechnen Sie die empirische Verteilungsfunktion an der Stelle 5, d.h. $F_Z(5)$.

12. Welcher Typ von Kardinalskalen wird jeweils in den folgenden Aussagen charakterisiert?

    (a) Die Merkmalsausprägungen entstehen durch Zählen.

    (b) Eine Merkmalsausprägung ist doppelt so groß wie eine andere.

    (c) Die Abstände zwischen je zwei Merkmalsausprägungen sind vergleichbar, aber es gibt keinen natürlichen Nullpunkt.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Es wird die Anzahl der Blitzeinschläge in Deutschland pro km^2 betrachtet. Die Gemeinde G gibt an, mit durchschnittlich 1.5 Einschlägen deutlich hinter dem deutschlandweiten Schnitt zu liegen.
Die Zufallsvariable

X : „Anzahl der Blitzeinschläge in Gemeinde G pro Jahr und km^2"

ist Poisson-verteilt mit Parameter $\lambda = 1.5$, außerdem ist bekannt: $\mu_X = \sigma_X^2 = \lambda$.

1. Berechnen Sie

   a) $P(X > 5)$

   b) $P(X \leq \mu_X)$

   c) $P(X^2 < \sigma_X^2)$

2. Bestimmen und interpretieren Sie für die Poisson-Verteilung mit Parameter $\lambda = 1.5$ ein gebräuchliches Schiefemaß.

In einem 1 km^2 großen Areal von Gemeinde G wurden die Blitzeinschläge der letzten 127 Jahre erfasst. Insgesamt schlugen in dieser Zeit $\sum_{i=1}^{127} x_i = 196$ Blitze ein.

3. Leiten Sie einen Schätzer für λ nach der Methode der Momente her und wenden Sie diesen auf die gegebene Stichprobe an. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 1.5.)

Testen Sie die Vermutung, dass $\lambda > 1.5$ ist. Hierzu formulieren Sie folgendes Hypothesenpaar:

$$H_0 : \lambda \leq \lambda_0 = 1.5 \quad \text{gegen} \quad H_1 : \lambda > \lambda_0.$$

Die hinreichend große Stichprobe erlaubt es Ihnen, den Test mittels approximativer Entscheidungsregel durchzuführen. Die dabei verwendete Testgröße ist approximativ standardnormalverteilt.

4. Berechnen Sie den Wert der realisierten Prüfgröße des approximativen Tests. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 0.4.)

5. Berechnen Sie den approximativen p-Wert für den beschriebenen Hypothesentest. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden einen p-Wert von 0.3.)

6. Treffen Sie anhand des p-Werts eine Testentscheidung bei einer Irrtumswahrscheinlichkeit von 5%.

7. Bestimmen Sie die kritische Schranke bei einer Irrtumswahrscheinlichkeit von 5%.

Schmierpapier zu Aufgabe 2

Aufgabe 3

1. Erläutern Sie <u>kurz</u> die Idee der Maximum-Likelihood-Schätzung für diskrete Stichprobenvariablen.

2. Der Fehler 2. Art beschreibt das Ereignis, die Nullhypothese ________________________, obwohl sie _______________ ist.

Christian und Nina machen eine Ski-Tagesfahrt. Da Nina langsamer fährt, wartet Christian nach jeder Abfahrt am Lift auf sie. Sei die Zufallsvariable

X : „Christians Wartezeit in Minuten"

exponentialverteilt mit unbekanntem Parameter $\lambda > 0$.

3. Geben Sie die Varianz von X in Abhängigkeit von λ an.

4. Sei in dieser Teilaufgabe $\lambda = 2$. Mit welcher Wahrscheinlichkeit wartet Christian

   a) mehr als 30 Sekunden, aber weniger als 90 Sekunden,

   b) mindestens 3 Minuten?

5. Zeigen Sie, dass die logarithmierte Likelihoodfunktion der Exponentialverteilung folgende Form hat:

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

6. Zeigen Sie, dass der Maximum-Likelihood-Schätzer für den Parameter λ gleich $1/\overline{X}_n$ ist. Gehen Sie davon aus, dass die hinreichende Bedingung für das Maximum erfüllt ist.

Beim Mittagessen auf der Kristall-Hütte äußert Nina die Befürchtung, dass Christian wertvolle Fahrzeit verliert, da er so lange auf sie warten muss. Um sie zu beruhigen, stoppt Christian bei den folgenden 10 Fahrten die Wartezeiten in Minuten:

Messung i	1	2	3	4	5	6	7	8	9	10
Zeit (x_i)	0.34	1.06	0.15	2.53	1.18	0.66	1.56	4.12	4.39	1.58

7. Geben Sie mit Hilfe von Teilaufgabe 6 den Schätzwert für den Erwartungswert von X an.

8. Wie heißt die Eigenschaft des Maximum-Likelihood-Schätzers, die besagt, dass

$$\widehat{h(\lambda)}_{ML} = h(\hat{\lambda}_{ML})$$

für eine beliebige Funktion $h(\cdot)$?

Am Ende des Skitages denkt Christian, dass er im Mittel 2 Minuten auf Nina warten muss. Nina hingegen vermutet eine mittlere Wartezeit, die auf jeden Fall von 2 Minuten abweicht.

Daher testet Nina ihre Vermutung auf dem 95%- Niveau:

$$H_0 : \mu_X = 2 \qquad \text{gegen} \qquad H_1 : \mu_X \neq 2 \tag{1}$$

Helfen Sie Nina, indem Sie die folgenden Schritte bearbeiten:

9. Zeigen Sie die Äquivalenz von (1) und dem Hypothesenpaar

$$H_0 : \lambda = 0.5 \qquad \text{gegen} \qquad H_1 : \lambda \neq 0.5 \tag{2}$$

Betrachten Sie folgende Stichprobenfunktion:

$$T_n(X_1, ..., X_n; \lambda) = 2\lambda \sum_{i=1}^{n} X_i \sim \chi^2(2n)$$

10. Berechnen Sie für den unbekannten Parameterwert λ zum Niveau $1 - \alpha = 95\%$ das zweiseitige realisierte Konfidenzintervall,
(Hinweis: $P\left(\frac{\chi^2_{\alpha/2;2n}}{2n\overline{X}_n} \leq \lambda \leq \frac{\chi^2_{1-\alpha/2;2n}}{2n\overline{X}_n}\right) = 1 - \alpha$)

und interpretieren Sie dieses <u>kurz</u>:

11. Treffen Sie eine Testentscheidung bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ und begründen Sie diese anhand geeigneter Werte, wenn Sie als Teststatistik $T_n(X_1, ..., X_n; \frac{1}{2})$ verwenden.

Schmierpapier zu Aufgabe 3

Aufgabe 4

In der Vergangenheit wurde festgestellt, dass die Niederschlagsmenge im Februar in Nürnberg gut durch eine normalverteilte Zufallsvariable beschrieben werden kann. Sei

X : „Niederschlagsmenge in Nürnberg im Februar (in mm)"

normalverteilt, d.h. $X \sim N(\mu_X = 34.1, \sigma_X^2 = 5)$.

1. Weshalb könnte es Probleme bei der Modellierung von Niederschlagsmengen durch die Normalverteilung geben? (Kurze Erläuterung)

2. Wie hoch ist die Wahrscheinlichkeit, dass die Niederschlagsmenge im Februar in Nürnberg

   a) genau 34 mm,

   b) mehr als 32.0875 mm, jedoch weniger als 36.1125 mm beträgt?

3. Geben Sie die Parameter für X an, wenn der Niederschlag in Metern statt in Millimetern gemessen wird.

Nun soll die Niederschlagsmenge in der Region Nürnberg-Fürth für den Februar modelliert werden. Dazu sei bekannt, dass die Niederschlagsmenge in Fürth im Februar (in mm), genannt Y, ebenfalls normalverteilt ist – es gelte $Y \sim N(\mu_Y = 39.9, \sigma_Y^2 = 10)$. Nehmen Sie zunächst an, die Zufallsvariablen X und Y wären stochastisch unabhängig.

4. Aufgrund welcher Eigenschaft der Normalverteilung ist die Zufallsvariable $(X + Y)$ wieder normalverteilt?

5. Berechnen Sie die Varianz von $X + Y$.

Die Zufallsvariablen Y und X sind nicht stochastisch unabhängig. Gehen Sie im Folgenden von $E[XY] = 1365$ aus.

6. Kann man allgemein von stochastischer Abhängigkeit immer auf Korreliertheit schließen? (Kurze Begründung)

7. Berechnen Sie den Korrelationskoeffizienten $\rho_{X,Y}$ für die Zufallsvariablen X und Y.

Verwenden Sie im weiteren Verlauf der Aufgabe den Wert $\rho_{X,Y} = 0.5$.

8. Berechnen Sie die Varianz für $X + Y$, die sich ergibt, wenn X und Y korreliert sind.

Seien nun (X, Y) gemeinsam bivariat normalverteilt.

9. Geben Sie alle nötigen Parameter für die bivariate Dichtefunktion von (X, Y) an (keine Rechnung nötig):

$$f_{(X,Y)}(x, y; \mu_X = \_\_\_\_\_, \mu_Y = \_\_\_\_\_, \sigma_X^2 = \_\_\_\_\_, \sigma_Y^2 = \_\_\_\_\_, \rho_{X,Y} = \_\_\_\_\_)$$

Die Messstation der Niederschlagsmenge in Fürth ist ausgefallen. Sie möchten nun die Verteilung dieser Niederschlagsmenge unter Verwendung des in Nürnberg bestimmten Wertes ableiten – dies ist mit der Verteilung $Y(x)$ möglich.
Es sei nun bekannt, dass die Verteilungsfunktion von $Y(x)$ gleich einer univariaten Normalverteilung ist mit den Parametern

$$Y(x) \sim N\left(\mu_{Y(x)} = \ \mu_Y + \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\ ,\ \sigma^2_{Y(x)} = \ \sigma_Y^2(1 - \rho_{X,Y})\right).$$

10. Wie wirkt sich ein überdurchschnittlich hoher beobachteter Wert x auf den Mittelwert der Zufallsvariablen $Y(x)$ aus? (Kurze Erläuterung)

11. Berechnen Sie die beiden Verteilungsparameter von $Y(x)$, wenn Sie wissen, dass die Niederschlagsmenge in Nürnberg $x = 35.1$ mm betragen hat:

$$Y(x = 35.1) \sim N\left(\mu_{Y(x=35.1)} = \_\_\_\_\_\_\_\_, \sigma^2_{Y(x=35.1)} = \_\_\_\_\_\_\_\_\right)$$

Verwenden Sie im weiteren Verlauf der Aufgabe die Werte $\mu_{Y(x=35.1)} = 41$ ***und*** $\sigma^2_{Y(x=35.1)} = 4$**.**

12. Berechnen Sie folgende Wahrscheinlichkeiten:

    a) $P(Y \leq 39.2)$

    b) $P(Y(x = 35.1) \leq 39.2)$

Hinweis: *Wenn Sie Teilaufgabe 12 nicht lösen konnten, gehen Sie von* $P(Y \leq 39.2) = 0.5$ *und* $P(Y(x = 35.1) \leq 39.2) = 0.35$ *aus.*

13. Begründen Sie kurz, wie der Unterschied der Ergebnisse in Teilaufgabe 12 zustande kommt.

Schmierpapier zu Aufgabe 4

Lösung Klausur WS13/14 (7.5ECTS)

Hinweis

Die Lösungen für die Aufgaben 1 bis 3 können Sie der Musterlösung der 10 ECTS Klausur entnehmen.

Aufgabe 4

1. Die Normalverteilung ist eine Verteilung auf ganz $\mathbb{R}$, und kann daher auch negative Werte annehmen. (0.5P)
2. (a) 0, da Punktwahrscheinlichkeit bei stetiger ZV (0.5P)

   (b) 0.6318 (0.5P Umformung, 0.5P Ergebnis)
3. $\mu_{\text{neu}} = \frac{\mu_X}{1000}$, $\sigma^2_{\text{neu}} = \frac{\sigma_X^2}{1000^2}$ (je 0.5P)
4. Reproduktivität (0.5P)
5. $\text{Var}(X+Y) \overset{\text{unabh.}}{=} \text{Var}(X) + \text{Var}(Y) = 15$ (0.5P)
6. Nein, da Korrelation nur **lineare** Abhängigkeit misst. (0.5P)
7. $\rho_{X,Y} = 0.6237$ (0.5P)
8. $\text{Var}(X+Y) \overset{\text{n. unabh.}}{=} \text{Var}(X) + \text{Var}(Y) + \text{Cov}(X,Y) = 23.82$
   (0.5P Ansatz, 0.5P Ergebnis)
9. $\mu_X = 34.1$, $\mu_Y = 39.9$, $\sigma_X^2 = 5$, $\sigma_Y^2 = 10$, $\rho_{X,Y} = 0.5$ (0.5P)
10. Er wirkt sich positiv aus, da $Y(x)$ Linearform – überdurchschnittlicher Wert. von x bedeutet, dass $x > \mu_X \leftrightarrow (x - \mu_X) > 0 \leftrightarrow$ positiver Steigungskoeffizient $\leftrightarrow$ im Mittel ist der Wert von $Y(x)$ höher (0.5P)
11. $\mu_{Y(35.1)} = 40.6071$, $\sigma^2_{Y(35.1)} = 5$ (je 0.5P)

12. (a) 0.4124 (0.5P)

    (b) 0.1841 (0.5P)

13. X und Y sind positiv korreliert (0.5P) (0.5P)
    – aus einem überdurchschnittlich hohen Wert von X ($x = 35.1 > 34.1 = \mu_X$) folgt mit höherer Wahrscheinlichkeit ein überdurchschnittlich hoher Wert von Y.

Klausur Statistik (10 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 12.02.2014 14:00 – 16:00 Uhr
Matrikelnummer			
Studienrichtung		**Sitzplatznummer**	
Semesterzahl		**Raum**	
Email (optional)			

Hinweise: Aufgabenblätter nicht auseinandertrennen!

Ergebnis:

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

Unterschrift des Kandidaten: ________________________________

Unterschrift des Prüfers: ____________________________________

Hilfsmittel: Zugelassen sind eine selbsterstellte maximal vierseitige (DIN A4 doppelseitig) Formelsammlung, die vom Lehrstuhl zum Download bereitgestellte Tabellensammlung ohne Eintragungen und die R-Reference-Card (by Jonathan Baron) ohne Eintragungen.
Darüber hinaus sind Taschenrechner zugelassen. Es dürfen jedoch keine Programme oder Programmteile verwendet werden, die nicht fest in den Taschenrechner eingebaut sind. Alle Hilfsmittel sind selbst mitzubringen.

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

In einer Umfrage zum Rauchverhalten wurden insgesamt 1000 Personen interviewt, die in Deutschland leben und mindestens 18 Jahre alt sind. Folgende Statistik zeigt den Anteil von Rauchern, Ex-Rauchern und Nie-Rauchern (Personen, die nie geraucht haben) aufgeteilt nach Geschlecht. Bekannt ist noch, dass $f(R|F) = 26.2\%$ ist.

Rauchverhalten → Geschlecht ↓	Raucher (R)	Ex-Raucher (E)	Nie-Raucher (N)	
Männer (M)				43.4%
Frauen (F)				
	30%	26.6%	43.4%	1

1. Berechnen Sie den Anteil der Raucherinnen. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 14.8%.)

2. Berechnen Sie $f(N \cap M)$, wenn Sie wissen, dass $f(E \cap M) = f(N \cap M)$ ist.

3. Wie viele der Befragten haben noch nie geraucht?

4. Interpretieren Sie $n(F \cap N) = 293$.

5. Um die Stärke der Abhängigkeit zwischen „Rauchverhalten" und „Geschlecht" zu untersuchen, wurde die mittlere quadratische Kontingenz berechnet. Diese beträgt 0.0379. Berechnen und interpretieren Sie das Cramérsche Kontingenzmaß V für die beiden Merkmale.

6. Nennen Sie ein geeignetes Streuungsmaß für das Merkmal „Rauchverhalten".

7. Nennen Sie ein geeignetes Lagemaß für das Merkmal „Geschlecht" und bestimmen Sie dieses.

8. Von den 1000 Untersuchungsteilnehmern werden nun 20 Personen zufällig ausgewählt und befragt (Doppelbefragungen seien zugelassen). Wie groß ist die Wahrscheinlichkeit, dass genau die Hälfte der 20 Befragten Raucher/innen sind, wenn der Prozentsatz der Raucher in der Grundgesamtheit genauso groß ist wie der Anteil der Raucher an den Untersuchungsteilnehmern?

9. Um welche Verteilung handelt es sich, wenn eine Person von den 1000 Personen befragt wird, ob er/sie raucht?

Folgende Daten geben den durchschnittlichen täglichen Zigarettenkonsum von 10 Personen an.

Person i	1	2	3	4	5	6	7	8	9	10
Anzahl der gerauchten Zigaretten pro Tag (Z)	1	3	4	8	8	15	21	21	10	7

10. Bestimmen Sie das untere Quartil der empirischen Verteilung von Z.

11. Berechnen Sie die empirische Verteilungsfunktion an der Stelle 5, d.h. $F_Z(5)$.

12. Welcher Typ von Kardinalskalen wird jeweils in den folgenden Aussagen charakterisiert?

    (a) Die Merkmalsausprägungen entstehen durch Zählen.

    (b) Eine Merkmalsausprägung ist doppelt so groß wie eine andere.

    (c) Die Abstände zwischen je zwei Merkmalsausprägungen sind vergleichbar, aber es gibt keinen natürlichen Nullpunkt.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Es wird die Anzahl der Blitzeinschläge in Deutschland pro km^2 betrachtet. Die Gemeinde G gibt an, mit durchschnittlich 1.5 Einschlägen deutlich hinter dem deutschlandweiten Schnitt zu liegen.
Die Zufallsvariable

X : „Anzahl der Blitzeinschläge in Gemeinde G pro Jahr und km^2"

ist Poisson-verteilt mit Parameter $\lambda = 1.5$, außerdem ist bekannt: $\mu_X = \sigma_X^2 = \lambda$.

1. Berechnen Sie

   a) $P(X > 5)$

   b) $P(X \leq \mu_X)$

   c) $P(X^2 < \sigma_X^2)$

2. Bestimmen und interpretieren Sie für die Poisson-Verteilung mit Parameter $\lambda = 1.5$ ein gebräuchliches Schiefemaß.

In einem 1 km^2 großen Areal von Gemeinde G wurden die Blitzeinschläge der letzten 127 Jahre erfasst. Insgesamt schlugen in dieser Zeit $\sum_{i=1}^{127} x_i = 196$ Blitze ein.

3. Leiten Sie einen Schätzer für λ nach der Methode der Momente her und wenden Sie diesen auf die gegebene Stichprobe an. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 1.5.)

Testen Sie die Vermutung, dass $\lambda > 1.5$ ist. Hierzu formulieren Sie folgendes Hypothesenpaar:

$$H_0 : \lambda \leq \lambda_0 = 1.5 \quad \text{gegen} \quad H_1 : \lambda > \lambda_0.$$

Die hinreichend große Stichprobe erlaubt es Ihnen, den Test mittels approximativer Entscheidungsregel durchzuführen. Die dabei verwendete Testgröße ist approximativ standardnormalverteilt.

4. Berechnen Sie den Wert der realisierten Prüfgröße des approximativen Tests. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden den Wert 0.4.)

5. Berechnen Sie den approximativen p-Wert für den beschriebenen Hypothesentest. (Wenn Sie diese Aufgabe nicht lösen können, verwenden Sie im Folgenden einen p-Wert von 0.3.)

6. Treffen Sie anhand des p-Werts eine Testentscheidung bei einer Irrtumswahrscheinlichkeit von 5%.

7. Bestimmen Sie die kritische Schranke bei einer Irrtumswahrscheinlichkeit von 5%.

Schmierpapier zu Aufgabe 2

Aufgabe 3

1. Erläutern Sie <u>kurz</u> die Idee der Maximum-Likelihood-Schätzung für diskrete Stichprobenvariablen.

2. Der Fehler 2. Art beschreibt das Ereignis, die Nullhypothese ________________________, obwohl sie ______________ ist.

Christian und Nina machen eine Ski-Tagesfahrt. Da Nina langsamer fährt, wartet Christian nach jeder Abfahrt am Lift auf sie. Sei die Zufallsvariable

X : „Christians Wartezeit in Minuten"

exponentialverteilt mit unbekanntem Parameter $\lambda > 0$.

3. Geben Sie die Varianz von X in Abhängigkeit von λ an.

4. Sei in dieser Teilaufgabe $\lambda = 2$. Mit welcher Wahrscheinlichkeit wartet Christian

   a) mehr als 30 Sekunden, aber weniger als 90 Sekunden,

   b) mindestens 3 Minuten?

5. Zeigen Sie, dass die logarithmierte Likelihoodfunktion der Exponentialverteilung folgende Form hat:

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

6. Zeigen Sie, dass der Maximum-Likelihood-Schätzer für den Parameter λ gleich $1/\overline{X}_n$ ist. Gehen Sie davon aus, dass die hinreichende Bedingung für das Maximum erfüllt ist.

Beim Mittagessen auf der Kristall-Hütte äußert Nina die Befürchtung, dass Christian wertvolle Fahrzeit verliert, da er so lange auf sie warten muss. Um sie zu beruhigen, stoppt Christian bei den folgenden 10 Fahrten die Wartezeiten in Minuten:

Messung i	1	2	3	4	5	6	7	8	9	10
Zeit (x_i)	0.34	1.06	0.15	2.53	1.18	0.66	1.56	4.12	4.39	1.58

7. Geben Sie mit Hilfe von Teilaufgabe 6 den Schätzwert für den Erwartungswert von X an.

8. Wie heißt die Eigenschaft des Maximum-Likelihood-Schätzers, die besagt, dass

$$\widehat{h(\lambda)}_{ML} = h(\hat{\lambda}_{ML})$$

für eine beliebige Funktion $h(\cdot)$?

Am Ende des Skitages denkt Christian, dass er im Mittel 2 Minuten auf Nina warten muss. Nina hingegen vermutet eine mittlere Wartezeit, die auf jeden Fall von 2 Minuten abweicht.

Daher testet Nina ihre Vermutung auf dem 95%- Niveau:

$$H_0 : \mu_X = 2 \qquad \text{gegen} \qquad H_1 : \mu_X \neq 2 \tag{1}$$

Helfen Sie Nina, indem Sie die folgenden Schritte bearbeiten:

9. Zeigen Sie die Äquivalenz von (1) und dem Hypothesenpaar

$$H_0 : \lambda = 0.5 \qquad \text{gegen} \qquad H_1 : \lambda \neq 0.5 \tag{2}$$

Betrachten Sie folgende Stichprobenfunktion:

$$T_n(X_1, ..., X_n; \lambda) = 2\lambda \sum_{i=1}^{n} X_i \sim \chi^2(2n)$$

10. Berechnen Sie für den unbekannten Parameterwert λ zum Niveau $1 - \alpha = 95\%$ das zweiseitige realisierte Konfidenzintervall,
(Hinweis: $P\left(\frac{\chi^2_{\alpha/2;2n}}{2n\overline{X}_n} \leq \lambda \leq \frac{\chi^2_{1-\alpha/2;2n}}{2n\overline{X}_n}\right) = 1 - \alpha$)

und interpretieren Sie dieses <u>kurz</u>:

11. Treffen Sie eine Testentscheidung bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ und begründen Sie diese anhand geeigneter Werte, wenn Sie als Teststatistik $T_n(X_1, ..., X_n; \frac{1}{2})$ verwenden.

Schmierpapier zu Aufgabe 3

Aufgabe 4

Ein Online-Versandhändler, der ausschließlich Schuhe und Bücher verkauft, möchte das Bestell- und Rücksendeverhalten seiner Kunden analysieren. Von 1000 Bestellungen liegen im `R`-Data Frame `amando` jeweils die folgenden Merkmale vor.

Merkmal *Art*:	Art der bestellten Artikel	Spalte `Art`
Merkmal *Zahlungsmethode*:	Gewählte Zahlungsmethode	Spalte `Zahlung`
Merkmal *Bestellwert*:	Wert der bestellten Artikel (in €)	Spalte `Bestellwert`
Merkmal *Retourenwert*:	Wert der bestellten Artikel (in €), die vom Kunden wieder zurückgeschickt wurden	Spalte `Retourenwert`

1. Geben Sie einen geeigneten Datentyp für die Spalte `Bestellwert` an.

2. Geben Sie den `R`-Befehl an, um den Gesamtwert der 1000 Bestellungen zu berechnen.

3. Geben Sie den `R`-Befehl an, um einen Vektor mit der wertmäßigen Rücksendequote der Bestellungen zu berechnen, d.h. mit dem jeweiligen Anteil des Retourenwerts am Bestellwert.

4. Vervollständigen Sie die folgende `R`-Funktion `kov` so, dass damit die Kovarianz für zwei Datenvektoren `vek1` und `vek2` berechnet wird.

```
kov=function(              , vek2){
  mean(          * vek2) - mean(vek1)*
}
```

5. Nennen Sie die Spalten des Data Frame `amando`, auf die die obige Funktion `kov` sinnvoll angewendet werden kann.

Betrachten Sie den zum Data Frame zugehörigen Output

```
> fkt1=function(y){
+   mean(y^2)-(mean(y)^2)/length(y)}
> fkt2=function(y){
+   mean(sum(y^2))-(sum(y)/mean(y))^2}
> fkt3=function(y){
+   (length(y)-1)/length(y)*var(y)}
> addmargins(table(amando$Zahlung,amando$Art)/nrow(amando))

              nurBuecher nurSchuhe Schuhe und Buecher  Sum
  Kreditkarte       0.25      0.20               0.15 0.60
  Rechnung          0.25      0.10               0.05 0.40
  Sum               0.50      0.30               0.20 1.00
> SB=amando[amando$Art=="Schuhe und Buecher","Zahlung"]
> Al=amando$Bestellwert
> Re.b=amando[amando$Zahlung=="Rechnung","Bestellwert"]
> Kr.b=amando[amando$Zahlung=="Kreditkarte","Bestellwert"]
> c(mean(Re.b),mean(Kr.b),mean(Al))
[1]  66.24760 116.03998  96.12303
> c(fkt1(Al),fkt2(Al),fkt3(Al))
[1]   10872.94 9882176.64    1642.54
```

6. Vervollständigen Sie die folgenden Aussagen:

(a) ______ % aller Bestellungen enthielten nur Schuhe .

(b) ______ % der Bestellungen enthielten nur Bücher und wurden per Kreditkarte bezahlt.

(c) Der durchschnittliche Bestellwert der Kreditkarten-Bestellungen beträgt ________ €.

(d) Die Varianz des Merkmals Bestellwert beträgt ____________ €^{2}.

7. Was wird mit dem Befehl

```
> table(SB)/length(SB)
```

erstellt?

Es wird angenommen, dass es sich bei den Rechnungs- und Kreditkarten-Bestellungen im Datensatz jeweils um einfache Zufallsstichproben handelt und dass für die Zufallsvariablen X: „Retourenwert bei Zahlung per Rechnung“ und Y: „Retourenwert bei Zahlung per Kreditkarte“ gelte

$$X \sim N(\mu_X, \sigma^2), \quad Y \sim N(\mu_Y, \sigma^2)$$

8. Bei einer Irrtumswahrscheinlichkeit von 5% soll die Vermutung getestet werden, dass der Retourenwert bei Zahlung per Rechnung im Mittel höher ist als der bei Zahlung per Kreditkarte. Vervollständigen Sie ausgehend von

```
> Re.r=amando[amando$Zahlung=="Rechnung","Retourenwert"]
> Kr.r=amando[amando$Zahlung=="Kreditkarte","Retourenwert"]
```

den folgenden Befehl, um einen geeigneten Test durchzuführen.

```
> t.test(Re.r,          ,                ,                              ,
        paired=FALSE, var.equal=TRUE, conf.level=0.95)
```

9. Die Prüfgröße des obigen Testes ist unter H_0 t-verteilt mit 998 Freiheitsgraden. Geben Sie den `R`-Befehl an, um den p-Wert des Testes zu bestimmen, wenn die realisierte Prüfgröße -6.7396 beträgt.

Schmierpapier zu Aufgabe 4

Lösung Klausur WS13/14 (10ECTS)

Aufgabe 1

1. $0.262(1 - 0.434) = 0.1483$ (1P)

2. $f(N \cap M) = \frac{1}{2}[0.434 - (0.3 - 0.1483)] = 0.141$ (1P)

3. $1000 \cdot 0.434 = 434$ (0.5P)

4. Es gibt 293 Nie-Raucherinnen. (0.5P)

5. $V = \sqrt{0.0379/\min\{3-1, 2-1\}} = 0.1947$ (1P)
   Die Assoziation/Zusammenhang zwischen den Merkmalen ist relativ schwach. (0.5P)

6. Entropie (0.5P)

7. Modus, Frauen (1P)

8. $f_{Bin}(10; 20, 0.3) = 0.0308$ (1P)

9. Bernoulli-Verteilung (0.5P)

10. 4 (0.5P)

11. 30% (0.5P)

12. (a) Absolutskala (0.5P)
    (b) Verhältnisskala oder Absolutskala (0.5P)
    (c) Intervallskala, Verhältnisskala oder Absolutskala (0.5P)

Aufgabe 2

1. a) $1 - F_{\text{Poi}}(5; 1.5) = 1 - 0.9955 = 0.0045$ (1P)

   b) $P(X \leq 1.5) = F_{\text{Poi}}(1; 1.5) = 0.5578$ (1P)

   c) $P(-1.2247 < X < 1.2247) = P(X < 1.2247)$
   $= F_{\text{Poi}}(1; 1.5) = 0.5578$ (1P)

2. Wenn z.B. der zweite Pearsonsche Schiefekoeffizient betrachtet wird:

   Bestimmung der fehlenden Größen (0.5P)
   Median: 1

   Berechnung des Schiefemaßes (0.5P)
   $(1.5 - 1)/\sqrt{1.5} = 0.4082$

   Richtige Interpretation des eben bestimmten Wertes (0.5P)
   $0.4082 > 0$: Verteilung ist rechtsschief

3. Herleitung (0.5P)
   z.B. $E[X] = \lambda \Rightarrow \hat{\lambda}_{\text{MM}} = \overline{X}_n$

   Berechnung: $\hat{\lambda}_{\text{MM}} = 196/127 = 1.5433$ (1P)

4. $\sqrt{n}\frac{\overline{x}_n - \lambda_0}{\sqrt{\lambda_0}} = \sqrt{127}\frac{1.5433-1.5}{\sqrt{1.5}} = 0.3984$ (1.5P)

5. $p = P(T_n > t_n | H_0) = 1 - \Phi(0.40) = 1 - 0.6554 = 0.3446$ (1P)

6. Vergleich des p-Werts mit dem Referenzwert (0.5P)

   $0.3446 > 0.05$

   Korrekte Testentscheidung basierend auf dem eben aufgestellten Vergleich (0.5P)

   Entweder:

   p-Wert < Referenzwert $\Rightarrow$ Ablehnung der Nullhypothese

   Oder:

   p-Wert > Referenzwert $\Rightarrow$ Nicht-Ablehnung der Nullhypothese

7. $\lambda_{0.95} = 1.6448.$ (0.5P)

Aufgabe 3

1. Suche nach Parameterwert $\hat{\theta}$, für den die Wahrscheinlichkeit für das Auftreten der realisierten Stichprobe maximal ist. (0.5P)

2. nicht abzulehnen; falsch (1P)

3. $Var(X) = \sigma_X^2 = \frac{1}{\lambda^2}$ (0.5P)

4. a) $P(0.5 < X < 1.5) = F_{Exp(2)}(1.5) - F_{Exp(2)}(0.5)$ (1P)
   $$= (1 - e^{-2\cdot 1.5}) - (1 - e^{-2\cdot 0.5}) = 0.3181$$
   b) $P(X \geq 3) = 1 - P(X \leq 3) = 1 - 0.9975 = 0.00248$ (0.5P)

5. $L(\lambda) = \prod\limits_{i=1}^{n} f_X(x;\lambda) = \prod\limits_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum\limits_{i=1}^{n} x_i}$ (1P)
   $\ln L(\lambda) = \ln\left(\lambda^n e^{-\lambda \sum\limits_{i=1}^{n} x_i}\right) = n\ln(\lambda) - \lambda \sum\limits_{i=1}^{n} x_i$ (0.5P)

6. $\frac{\partial}{\partial\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum\limits_{i=1}^{n} x_i \overset{!}{=} 0 \quad \Longrightarrow \hat{\lambda} = \frac{n}{\sum\limits_{i=1}^{n} x_i} = \frac{1}{\bar{X}_n} = \left(\bar{X}_n\right)^{-1}$ (1P+0.5P)

7. $\hat{\mu}_X = \frac{1}{\hat{\lambda}} = \frac{1}{\frac{n}{\sum\limits_{i=1}^{n} x_i}} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = 1.757$ (0.5P)

8. Invarianzeigenschaft (0.5P)

9. $\mu_X = \frac{1}{\lambda} \quad \Rightarrow \quad \mu_X = 2 \Leftrightarrow \frac{1}{\lambda} = 2 \Leftrightarrow \lambda = \frac{1}{2}$ (0.5P)

10. $\left[\frac{9.59}{2\cdot 17.57} \quad ; \quad \frac{34.17}{2\cdot 17.57}\right] = [0.2729 \quad ; \quad 0.9724]$ (0.5P)

    Endweder: (0.5P)
    Mit einer Vertrauenswürdigkeit von 95% liegt der Parameter λ in diesem Intervall
    oder:
    In 95% der Fälle liegt der Parameter λ in diesem Intervall

11. **Endweder:**

    $\chi^2_{0.025;20} = 9.59 < t_{10}(x_1, ..., x_{10}; 2) = 17.57 < 34.17 = \chi^2_{0.975;20}$ (1P)

    $\Rightarrow$ H_0 nicht ablehnen auf 5%-Niveau.

    oder:

    $\lambda_0 \in KI \quad \Rightarrow \quad H_0$ nicht ablehnen auf 5%-Niveau.

Aufgabe 4

1. numeric (*oder* integer) (0.5P)

2. z.B. `>sum(amando$Bestellwert)` (1P)

3. z.B. `>amando$Retourenwert/amando$Bestellwert` (1P)

4. (1.5P)
```
kov=function(vek1, vek2){
  mean(vek1*vek2)-mean(vek1)*mean(vek2)
}
```

5. Bestellwert und Retourenwert (0.5P)

6. (a) 30 (0.5P)

   (b) 25 (0.5P)

   (c) 116.03998 (0.5P)

   (d) 1642.54 (0.5P)

7. Tabelle der relativen Häufigkeiten der Zahlungsmethode von Bestellungen, die Schuhe und Bücher enthielten (1P)

8. (1.5P)
```
> t.test(Re.r,Kr.r,mu=0,alternative="greater",
        paired=FALSE, var.equal=TRUE, conf.level=0.95)
```

9. `> 1-pt(-6.7396,998)` (1P)