Klausur Statistik (7.5 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Dienstag, 27.02.2018
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.
- 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

In einer fränkischen Stadt leben 25.000 Menschen. Die Zufallsvariable

$$Z\text{: "Jährliches Bruttoeinkommen eines Einwohners [Euro]"}$$

sei von Person zu Person stochastisch unabhängig und identisch verteilt. Außerdem sei Z normalverteilt mit Varianz $\sigma_Z^2 = 5.000$. Sie interessieren sich für das mittlere jährliche Bruttoeinkommen eines Einwohners. Hierzu stellen Sie das folgende Hypothesenpaar auf:

$$H_0 : \mu_Z = \mu_0 = 40.000 \quad H_1 : \mu_Z \neq \mu_0 = 40.000$$

und legen eine Irrtumswahrscheinlichkeit von 1% fest.

1. Geben Sie die theoretische Prüfgröße sowie deren Erwartungswert und Varianz unter H_0 an.

2. Wie lautet der Ablehnungsbereich? (mit konkretem Wert)

In einer Umfrage wird das jährliche Bruttoeinkommen von 100 Einwohnern erhoben. Hierbei ergibt sich $\bar{z}_{100} = 40.050$.

3. Berechnen Sie die Varianz des Stichprobenmittels.

4. Berechnen Sie die realisierte Prüfgröße.

5. Treffen Sie mit den vorliegenden Informationen und kurzer Begründung eine Testentscheidung. (Falls Sie 2. und 4. nicht lösen konnten, verwenden Sie als Ersatzwert für die kritische Schranke 1.96 und als Ersatzwert für die realisierte Prüfgröße 6.07)

6. Ist der p-Wert größer oder kleiner als die Irrtumswahrscheinlichkeit α? (kurze Begründung, keine Berechnung)

7. Berechnen Sie die Güte des Tests an der Stelle $\mu = 40.005$.

8. Nennen Sie eine Möglichkeit, wie die Güte eines Tests erhöht werden kann.

9. Wie groß ist die Wahrscheinlichkeit, dass aus den 100 Einwohnern einer ungerundet genau 40.000 Euro jährliches Bruttoeinkommen erhält?

10. Nun sei $\mu_Z = 40.000$. Wie und mit welchen Parametern ist die Zufallsvariable G: "Summe des jährlichen Bruttoeinkommens aus zwei Stadtbewohnern [Euro]" verteilt?

Ein Haushalt aus 2 Personen gelte als arm, wenn er weniger als 38.000 Euro jährliches Bruttoeinkommen bezieht. Verwenden Sie im Folgenden $F \sim N(38.300, 12.000)$ für ein Ehepaar, in dem nur eine Person berufstätig ist.

11. Wie hoch ist die Wahrscheinlichkeit, dass ein Ehepaar mit nur einer berufstätigen Person arm ist?

12. Geben Sie die Maximalstelle der Dichte der Verteilung von F an. Es ist hierbei keine Berechnung erforderlich.

13. Geben Sie an, gegen welchen Wert die Verteilungsfunktion von F konvergiert, wenn das Argument gegen unendlich konvergiert. Es ist hierbei keine Berechnung erforderlich.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

1. In einem Bistro werden unter anderem Cupcakes verkauft. Der Ladenbesitzer macht folgende Beobachtungen für die Zufallsvariable Y: "Anzahl an verkauften Cupcakes pro Stunde":

Anzahl verkaufter Cupcakes y	1	2	3	≥ 4
Wahrscheinlichkeit $f_Y(y)$	0.33	0.35	0.23	?

a) Welcher Wert fehlt in der Tabelle, damit eine Wahrscheinlichkeitsfunktion definiert ist?

b) Bestimmen Sie den Wert der Verteilungsfunktion von Y an der Stelle $y = 1$ und $y = 2.5$.

c) Wie groß ist die Wahrscheinlichkeit innerhalb einer Stunde 2 oder mehr Cupcakes zu verkaufen?

2. Außerdem möchte der Ladenbesitzer wissen, wie viele Kaffeebestellungen er pro Stunde erwarten kann und betrachtet hierfür die Zufallsvariable X: “Anzahl der Kaffeebestellungen pro Stunde”. Er geht davon aus, dass die Zufallsvariable X einer Poissonverteilung mit unbekanntem Parameter λ folgt und erhebt folgende einfache Stichprobe:

Stunde i	1	2	3	4	5
Anzahl Bestellungen x_i	4	5	2	6	3

Es gilt: $\sum_{i=1}^{5} x_i = 20$.

a) Zeigen Sie mittels der Methode der Momente, dass das Stichprobenmittel ein Schätzer für den Parameter λ ist und berechnen sie das realisierte Stichprobenmittel.

b) Bestimmen Sie den Erwartungswert des Stichprobenmittels.

3. Sie betrachten für einen Parameter ρ den Schätzer R_n. Welche Bedingung muss erfüllt sein, damit R_n ein erwartungstreuer Schätzer für ρ ist?

4. Sie möchten die Hypothese, dass der wahre Parameter λ den Wert 2 überschreitet mit einer Irrtumswahrscheinlichkeit von 4.87% testen. Dazu stellen Sie folgendes Hypothesenpaar auf

$$H_0 : \lambda \leq \lambda_0 = 2 \text{ vs. } H_1 : \lambda > 2$$

a) Geben Sie das 95.13% Quantil der Poissonverteilung mit Parameter $\lambda = 10$ an.

b) Treffen Sie mittels der exakten Entscheidungsregel eine Testentscheidung mit Hilfe des in Teilaufgabe a) bestimmten Quantils. Wenn Sie Teilaufgabe a) nicht lösen konnten, verwenden Sie 12 als Ersatzergebnis.

c) Wie groß ist die Güte des Tests, wenn die Wahrscheinlichkeit für den Fehler 2. Art 75% beträgt?

5. Im Folgenden sei $\lambda = 4$, d.h. für die Zufallsvariable X: “Anzahl der Kaffeebestellungen pro Stunde” gilt $X \sim Pois(4)$.

a) Berechnen Sie die Wahrscheinlichkeit, dass innerhalb einer Stunde mindestens 3 und höchstens 5 Kaffeebestellungen eingehen.

b) Berechnen Sie die Varianz von X.

c) Bestimmen Sie die Verteilung der Zufallsvariablen Y: “Anzahl der Kaffeebestellungen innerhalb von 2 Stunden”.

d) Es wird ein zweites Bistro eröffnet, dessen Anzahl an Kaffeebestellungen stochastisch unabhängig und identisch verteilt zu der Anzahl an Bestellungen im ersten Bistro ist. Wie ist die Zufallsvariable Z: “Anzahl der Kaffeebestellungen in beiden Läden innerhalb einer Stunde” verteilt?

6. Wartezeiten zwischen zwei Poissonereignissen sind exponentialverteilt. Welchen Wert nimmt der Parameter λ der Exponentialverteilung an, wenn im Durchschnitt 5 Poissonereignisse pro Zeiteinheit auftreten?

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

1. Sie beobachten die relative Veränderung r_i des jährlichen Bruttoinlandsprodukts in Deutschland in den Jahren 2008 bis 2016.

i	2008	2009	2010	2011	2012	2013	2014	2015	2016
r_i	0.02	-0.04	0.05	0.05	0.02	0.02	0.04	0.04	0.03

a) Geben Sie den Median der Beobachtungen an.

b) Wie ändert sich der Median, wenn die Beobachtung für das Jahr 2009 $r_{2009} = -0.05$ ist? (wird größer/wird kleiner/bleibt gleich)

c) Betrachten Sie die Wachstumsfaktoren $1 + r_i$ und berechnen Sie das geometrische Mittel.

d) Berechnen Sie die Stichprobenvarianz der Werte r_i.
Hinweis: $\sum_{i=1}^{9} r_i = 0.23, \quad \sum_{i=1}^{9} r_i^2 = 0.0119$

e) Nehmen Sie an, dass die Beobachtungen r_i Realisationen einer normalverteilten Zufallsvariable sind und die Realisation der Stichprobenvarianz 0.0006 ist. Testen Sie zum Konfidenzniveau von 95% die Nullhypothese, dass die Varianz der Grundgesamtheit größer oder gleich 0.001 ist:

$$H_0 : \sigma_0^2 \geq 0.001$$

Verwenden Sie dabei die Teststatistik:

$$T_n = (n-1)\frac{S_X^2}{\sigma_0^2}.$$

i) Wie ist die Prüfgröße unter H_0 verteilt? Wie lautet der bzw. lauten die Parameter der Verteilung?

ii) Geben Sie den kritischen Bereich mit konkreten Werten an.

iii) Wie lautet die Testentscheidung?

f) Bei einer Befragung gehen 70% der Personen davon aus, dass die relative Veränderung des Bruttoinlandsprodukts zum Vorjahr 2018 größer ist als 2016, 20% glauben sie wird den selben Wert annehmen, 10% vermuten einen Rückgang. Berechnen Sie die Gini-Entropie für diese Einschätzung.

2. In einem Onlineshop ist die Zeit t_i zwischen dem ersten Besuch eines Kunden i auf dem Internetauftritt und dem Kauf eines Produkts [in Minuten] gegeben:

i	1	2	3	4	5
t_i	13	7	29	5	11

Nehmen Sie an, dass diese Zeitdauern unabhängig und identisch verteilte Realisationen einer Exponentialverteilung mit Parameter λ sind. Die Log-Likelihood Funktion für dieses Modell ist damit gegeben durch:

$$LL(\lambda|t_1, \ldots, t_n) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} t_i$$

a) Wie lautet die Ableitung der Log-Likelihood Funktion bzgl. λ?

b) Zeigen Sie, dass der Maximum Likelihood Schätzer für λ durch $\widehat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} t_i}$ gegeben ist und berechnen Sie den konkreten Schätzwert für die Beobachtungen t_i. Hinweis: Gehen Sie davon aus, dass die hinreichenden Bedingungen für ein Maximum erfüllt sind.

Nehmen Sie im Folgenden an, dass den beschriebenen Wartezeiten bis zum Kauf eines Produkts tatsächlich eine Exponentialverteilung zu Grunde liegt und der Parameter $\lambda = 0.4$ ist. Die Dichtefunktion dieser Zufallsvariable X lautet $f(x) = \lambda \exp(-\lambda x)$, die Verteilungsfunktion $F(x) = 1 - \exp(-\lambda x)$.

c) Wie groß ist die Wahrscheinlichkeit, dass ein Kunde höchstens 4 Minuten nach dem erstem Besuch des Onlineshops einen Kauf tätigt?

d) Für Erwartungswert und Varianz von X gilt: $E(X) = \frac{1}{\lambda}$ und $Var(X) = \frac{1}{\lambda^2}$. Wie lautet das zweite Moment $E(X^2)$?

e) Es seien zwei exponentialverteilte Zufallsvariablen $X \sim Exp(\lambda = 3)$ und $Y \sim Exp(\lambda = 5)$ gegeben. Die Kovarianz zwischen X und Y betrage $Cov(X, Y) = 0.02$. Wie groß ist die Varianz von $X + Y$?

f) Berechnen Sie die Korrelation von X und Y.

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

Sie sind Besitzer eines 'Escape Room', in dem Personengruppen durch gemeinsames Rätsellösen einen Ausweg aus den von Ihnen gestalteten Räumlichkeiten suchen.

1. Um mehr über Ihre Teilnehmer zu erfahren erstellen Sie einen Fragebogen.

a) Sie wollen etwas über das durchschnittliche Alter (μ_J) Ihrer Teilnehmer herausfinden. Sie wissen, dass $\sigma_J^2 = 17.31$, haben jedoch keine Information über die Verteilung des Stichprobenmittelwertes. Wie viele Personen müssen Sie mindestens befragen, damit der Stichprobenmittelwert mit einer Irrtumswahrscheinlichkeit von 5% um höchstens 2 Jahre vom wahren Mittelwert abweicht?

Nehmen Sie nun an, dass das Alter Ihrer Teilnehmer (J) unabhängig und identisch normalverteilt ist. Sie haben 90 Personen befragt und ein durchschnittliches Alter von $\bar{j}_{90} = 25$ Jahren Ihrer Teilnehmer dieser Stichprobe kalkuliert.

b) Berechnen Sie das realisierte zweiseitige 75% Konfidenzintervall für das Durchschnittsalter Ihrer Teilnehmer. (Hinweis: $\sigma_J^2 = 17.31$)

c) Interpretieren Sie das realisierte Konfidenzintervall.

2. Sie bieten außerdem Getränke an. Sie haben in folgender Kontingenztabelle dargestellt, ob ein Teilnehmer mindestens 25 Jahre (≥ 25) oder jünger als 25 Jahre (< 25) alt war und ob der Teilnehmer ein alkoholisches A oder nicht-alkoholisches Getränk N bestellt hat.

	< 25	≥ 25	$\sum$
A		26	58
N	9	23	
$\sum$	41	49	90

a) Geben Sie n_{12} an und interpretieren Sie den Wert.

b) Berechnen Sie $f_{2\cdot}$ und interpretieren Sie den Wert.

Sie haben die mittlere quadratische Kontingenz berechnet: $\phi^2 = 0.0676$.

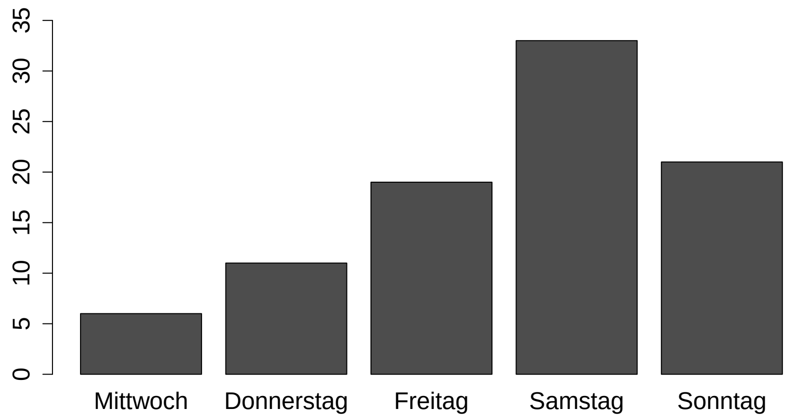
c) Berechnen Sie Cramér's V und interpretieren Sie Ihr Ergebnis.

Nun wollen Sie wissen, ob die Art eines bestellten Getränks unabhängig vom Alter des Bestellers ist und führen dazu einen χ^2-Unabhängigkeitstest mit einem 10% Signifikanzniveau durch. Als Wert der Teststatistik haben Sie $\chi^2 = 6.0826$ erhalten.

d) Geben Sie den Wert der kritischen Schranke für Ihren Test an.

e) Treffen Sie eine Testentscheidung und begründen Sie diese.

3. In Ihrem Fragebogen haben Sie zusätzlich noch den Tag erfragt, an dem der Teilnehmer in Ihrem Escape Room war. Das folgende Balkendiagramm zeigt Ihre Ergebnisse.



a) Welchen Merkmalstyp besitzt das Merkmal T: "Tag, an dem der Teilnehmer im Escape Room war"?

b) Warum kann das Merkmal T nicht ordinal skaliert sein?

c) Geben Sie den Modus des Merkmals T an.

4. Sie wissen, dass 15% aller Teilnehmer schon einmal in einem Escape Room waren. Sei nun B: "Teilnehmer war bereits vorher schon einmal in einem Escape Room" binomialverteilt. Es kommen heute 7 neue Teilnehmer zu Ihnen.

a) Wie hoch ist die Wahrscheinlichkeit, dass keiner der heutigen Teilnehmer schon vorher in einem Escape Room war?

b) Wie hoch ist die Wahrscheinlichkeit, dass höchstens 5 Teilnehmer vorher in einem Escape Room waren?

c) Ab wie vielen heutigen Teilnehmern kann $B_n \sim Bin(n, p = 0.15)$ durch eine Normalverteilung approximiert werden?

Schmierpapier zu Aufgabe 4

Lösung Klausur WiSe 17/18 (7.5 ECTS)

Lösung 1

1. $T_n = \sqrt{n}\frac{\bar{Z}_n - \mu_0}{\sigma} \qquad E[T_n] = 0, Var[T_n] = 1$ 1P

2. $|t_n| > \lambda_{1-\alpha/2} \qquad \lambda_{0.995} = 2.5758$ 1P

3. 50 0.5P

4. 7.0711 0.5P

5. $t_n > \lambda_{1-\alpha/2} = 2.5758$
   H_0 kann bei Irrtums-WS von $\alpha = 1\%$ abgelehnt werden. 1P

6. $p < \alpha$, da H_0 abgelehnt wird. 0.5P

7. $\pi = 2 - \Phi(1.87) - \Phi(3.28) = 0.0312$ 1P

8. Erhöhung von Stichprobenumfang n 0.5P

9. 0 0.5P

10. $N(80.000, 10.000)$ 1P

11. $\Phi(\frac{38.000-38.300}{\sqrt{12.000}}) = \Phi(-2.74) = 1 - \Phi(2.74) = 0.0031$ 1.5P

12. 38.300 0.5P

13. 1 0.5P

Lösung 2

1a) 0.09 — 0.5 P

1b) $F_Y(1) = 0.33$ und $F_Y(2.5) = 0.68$ — 1P

1c) $P(Y \geq 2) = 1 - P(Y \leq 1) = 0.67$ — 1P

2a) $\mathbb{E}[X] = \lambda \Leftrightarrow \hat{\lambda}_{MM} = M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i^1 = \bar{X}_n$ und $\hat{\lambda} = 4$ — 1P

2b) $E[\bar{X}_n] = E[X] = \lambda$ — 1P

3. $E[R_n] = \rho$ — 0.5 P

4a) 15 — 0.5 P

4b) $\sum_{i=1}^{5} X_i = 20 > 15 \Rightarrow$ Lehne H_0 bei einer Irrtumswahrscheinlichkeit von 4.87% ab. 1 P

4c) 25% — 0.5P

5a) $P(3 \leq X \leq 5) = 0.7851 - 0.2381 = 0.547/0.5471$ — 1P

5b) $Var(X) = 4$ — 0.5P

5c) $Y \sim Pois(8)$ — 0.5P

5d) $Z \sim Pois(8)$ — 0.5P

6. 5 — 0.5P

Lösung 3

1. a) Median: 0.03 — 0.5 P

   b) bleibt gleich — 0.5 P

   c) Geometrisches Mittel: 1.0252 — 0.5 P

   d) Stichprobenvarianz: 0.0008 — 1 P

   e) i.) Prüfgröße ist χ^2 verteilt mit Parameter $n-1=8$. — 0.5 P

   e) ii.) Lehne H_0 ab, wenn $t_n < \chi^2_{0.05;8} = 2.73$. — 1 P

   e) iii.) Lehne H_0 ab, wenn $t_n < \chi^2_{0.05;8} \Rightarrow 8 \cdot \frac{0.0006}{0.001} = 4.8 > 2.73$
   $\Rightarrow H_0$ beibehalten — 1 P

   f) Gini-Entropie: $0.7 \cdot 0.3 + 0.2 \cdot 0.8 + 0.1 \cdot 0.9 = 0.46$ — 0.5 P

2. a) $\frac{\partial LL(\lambda|t_1,\ldots,t_n)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n t_i$ — 1 P

   b) $\frac{\partial LL(\lambda|t_1,\ldots,t_n)}{\partial \lambda} = 0 \Leftrightarrow \widehat{\lambda}_{\mathrm{ML}} = \frac{n}{\sum_{i=1}^n t_i}, \quad \widehat{\lambda}_{\mathrm{ML}} = 0.0769$ — 1 P

   c) $P(X \leq 4) = 1 - \exp(-0.4 \cdot 4) = 0.7981$ — 0.5 P

   d) $E(X^2) = Var(X) + E(X)^2 = \frac{2}{\lambda^2}$ — 1 P

   e) $Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y) = 0.1911$ — 0.5 P

   f) $Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = 0.3$ — 0.5 P

Lösung 4

1. a) $n \geq \frac{\sigma^2}{\alpha\epsilon^2} = \frac{17.31}{0.05 * 2^2} = 86.5500$, d.h. es müssen mindestens 87 Personen befragt werden. 1 P

   b) $\lambda_{0.875} = 1.1503$, d.h. $[24.4955; 25.5045]$ 1 P

   c) Der unbekannte Stichprobenmittelwert liegt mit einer Vertrauenswürdigkeit von 75% im berechneten realisierten Intervall 0.5 P

2. a) $n_{12} = 26$, 0.5 P
   d.h., 26 Teilnehmer waren unter 25 und haben ein alkoholisches Getränk bestellt. 0.5 P

   b) $f_{2\cdot} = 32/90 = 0.3556$, 0.5 P
   d.h. 35.56% aller Teilnehmer haben ein nicht-alkoholisches Getränk bestellt. 0.5 P

   c) Cramér's V$= \sqrt{\phi^2/1} = 0.2600$, 0.5 P
   d.h., dass der Grad der Abhängigkeit der Merkmale 'Getränk' und 'Alter' eher gering ist. 0.5 P

   d) kritische Schranke $\chi^2(0.9; 1) = 2.706$ 0.5 P

   e) Testentscheidung: Da $6.0826 > 2.706$: 0.5 P
   Die Nullhypothese kann auf dem 10% Signifikanzniveau abgelehnt werden. 0.5 P

3. a) Merkmalstyp: qualitativ 0.5 P

   b) Das Merkmal T kann nicht ordinal skaliert sein, da keine Rangordnung der Merkmalsausprägungen gegeben ist. 0.5 P

   c) Modus: Samstag 0.5 P

4. a) $f_{Bin}(0; 7, 0.15) = 0.3206$ 0.5 P

   b) $F_{Bin}(5; 7, 0.15) = 0.9999$ 0.5 P

   c) $np(1-p) > 9$, d.h. $n > 70.5882$. Ab 71 heutigen Teilnehmern kann B_n durch eine Normalverteilung approximiert werden. 0.5 P

Klausur Statistik (10 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Dienstag, 27.02.2018
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.
- 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

In einer fränkischen Stadt leben 25.000 Menschen. Die Zufallsvariable

$$Z\text{: "Jährliches Bruttoeinkommen eines Einwohners [Euro]"}$$

sei von Person zu Person stochastisch unabhängig und identisch verteilt. Außerdem sei Z normalverteilt mit Varianz $\sigma_Z^2 = 5.000$. Sie interessieren sich für das mittlere jährliche Bruttoeinkommen eines Einwohners. Hierzu stellen Sie das folgende Hypothesenpaar auf:

$$H_0 : \mu_Z = \mu_0 = 40.000 \quad H_1 : \mu_Z \neq \mu_0 = 40.000$$

und legen eine Irrtumswahrscheinlichkeit von 1% fest.

1. Geben Sie die theoretische Prüfgröße sowie deren Erwartungswert und Varianz unter H_0 an.

2. Wie lautet der Ablehnungsbereich? (mit konkretem Wert)

In einer Umfrage wird das jährliche Bruttoeinkommen von 100 Einwohnern erhoben. Hierbei ergibt sich $\bar{z}_{100} = 40.050$.

3. Berechnen Sie die Varianz des Stichprobenmittels.

4. Berechnen Sie die realisierte Prüfgröße.

5. Treffen Sie mit den vorliegenden Informationen und kurzer Begründung eine Testentscheidung. (Falls Sie 2. und 4. nicht lösen konnten, verwenden Sie als Ersatzwert für die kritische Schranke 1.96 und als Ersatzwert für die realisierte Prüfgröße 6.07)

6. Ist der p-Wert größer oder kleiner als die Irrtumswahrscheinlichkeit α? (kurze Begründung, keine Berechnung)

7. Berechnen Sie die Güte des Tests an der Stelle $\mu = 40.005$.

8. Nennen Sie eine Möglichkeit, wie die Güte eines Tests erhöht werden kann.

9. Wie groß ist die Wahrscheinlichkeit, dass aus den 100 Einwohnern einer ungerundet genau 40.000 Euro jährliches Bruttoeinkommen erhält?

10. Nun sei $\mu_Z = 40.000$. Wie und mit welchen Parametern ist die Zufallsvariable G: "Summe des jährlichen Bruttoeinkommens aus zwei Stadtbewohnern [Euro]" verteilt?

Ein Haushalt aus 2 Personen gelte als arm, wenn er weniger als 38.000 Euro jährliches Bruttoeinkommen bezieht. Verwenden Sie im Folgenden $F \sim N(38.300, 12.000)$ für ein Ehepaar, in dem nur eine Person berufstätig ist.

11. Wie hoch ist die Wahrscheinlichkeit, dass ein Ehepaar mit nur einer berufstätigen Person arm ist?

12. Geben Sie die Maximalstelle der Dichte der Verteilung von F an. Es ist hierbei keine Berechnung erforderlich.

13. Geben Sie an, gegen welchen Wert die Verteilungsfunktion von F konvergiert, wenn das Argument gegen unendlich konvergiert. Es ist hierbei keine Berechnung erforderlich.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

1. In einem Bistro werden unter anderem Cupcakes verkauft. Der Ladenbesitzer macht folgende Beobachtungen für die Zufallsvariable Y: "Anzahl an verkauften Cupcakes pro Stunde":

Anzahl verkaufter Cupcakes y	1	2	3	≥ 4
Wahrscheinlichkeit $f_Y(y)$	0.33	0.35	0.23	?

a) Welcher Wert fehlt in der Tabelle, damit eine Wahrscheinlichkeitsfunktion definiert ist?

b) Bestimmen Sie den Wert der Verteilungsfunktion von Y an der Stelle $y = 1$ und $y = 2.5$.

c) Wie groß ist die Wahrscheinlichkeit innerhalb einer Stunde 2 oder mehr Cupcakes zu verkaufen?

2. Außerdem möchte der Ladenbesitzer wissen, wie viele Kaffeebestellungen er pro Stunde erwarten kann und betrachtet hierfür die Zufallsvariable X: "Anzahl der Kaffeebestellungen pro Stunde". Er geht davon aus, dass die Zufallsvariable X einer Poissonverteilung mit unbekanntem Parameter λ folgt und erhebt folgende einfache Stichprobe:

Stunde i	1	2	3	4	5
Anzahl Bestellungen x_i	4	5	2	6	3

Es gilt: $\sum_{i=1}^{5} x_i = 20$.

a) Zeigen Sie mittels der Methode der Momente, dass das Stichprobenmittel ein Schätzer für den Parameter λ ist und berechnen sie das realisierte Stichprobenmittel.

b) Bestimmen Sie den Erwartungswert des Stichprobenmittels.

3. Sie betrachten für einen Parameter ρ den Schätzer R_n. Welche Bedingung muss erfüllt sein, damit R_n ein erwartungstreuer Schätzer für ρ ist?

4. Sie möchten die Hypothese, dass der wahre Parameter λ den Wert 2 überschreitet mit einer Irrtumswahrscheinlichkeit von 4.87% testen. Dazu stellen Sie folgendes Hypothesenpaar auf

$$H_0 : \lambda \leq \lambda_0 = 2 \text{ vs. } H_1 : \lambda > 2$$

a) Geben Sie das 95.13% Quantil der Poissonverteilung mit Parameter $\lambda = 10$ an.

b) Treffen Sie mittels der exakten Entscheidungsregel eine Testentscheidung mit Hilfe des in Teilaufgabe a) bestimmten Quantils. Wenn Sie Teilaufgabe a) nicht lösen konnten, verwenden Sie 12 als Ersatzergebnis.

c) Wie groß ist die Güte des Tests, wenn die Wahrscheinlichkeit für den Fehler 2. Art 75% beträgt?

5. Im Folgenden sei $\lambda = 4$, d.h. für die Zufallsvariable X: "Anzahl der Kaffeebestellungen pro Stunde" gilt $X \sim Pois(4)$.

a) Berechnen Sie die Wahrscheinlichkeit, dass innerhalb einer Stunde mindestens 3 und höchstens 5 Kaffeebestellungen eingehen.

b) Berechnen Sie die Varianz von X.

c) Bestimmen Sie die Verteilung der Zufallsvariablen Y: “Anzahl der Kaffeebestellungen innerhalb von 2 Stunden”.

d) Es wird ein zweites Bistro eröffnet, dessen Anzahl an Kaffeebestellungen stochastisch unabhängig und identisch verteilt zu der Anzahl an Bestellungen im ersten Bistro ist. Wie ist die Zufallsvariable Z: “Anzahl der Kaffeebestellungen in beiden Läden innerhalb einer Stunde” verteilt?

6. Wartezeiten zwischen zwei Poissonereignissen sind exponentialverteilt. Welchen Wert nimmt der Parameter λ der Exponentialverteilung an, wenn im Durchschnitt 5 Poissonereignisse pro Zeiteinheit auftreten?

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

1. Sie beobachten die relative Veränderung r_i des jährlichen Bruttoinlandsprodukts in Deutschland in den Jahren 2008 bis 2016.

i	2008	2009	2010	2011	2012	2013	2014	2015	2016
r_i	0.02	−0.04	0.05	0.05	0.02	0.02	0.04	0.04	0.03

a) Geben Sie den Median der Beobachtungen an.

b) Wie ändert sich der Median, wenn die Beobachtung für das Jahr 2009 $r_{2009} = -0.05$ ist? (wird größer/wird kleiner/bleibt gleich)

c) Betrachten Sie die Wachstumsfaktoren $1 + r_i$ und berechnen Sie das geometrische Mittel.

d) Berechnen Sie die Stichprobenvarianz der Werte r_i.
Hinweis: $\sum_{i=1}^{9} r_i = 0.23, \quad \sum_{i=1}^{9} r_i^2 = 0.0119$

e) Nehmen Sie an, dass die Beobachtungen r_i Realisationen einer normalverteilten Zufallsvariable sind und die Realisation der Stichprobenvarianz 0.0006 ist. Testen Sie zum Konfidenzniveau von 95% die Nullhypothese, dass die Varianz der Grundgesamtheit größer oder gleich 0.001 ist:

$$H_0 : \sigma_0^2 \geq 0.001$$

Verwenden Sie dabei die Teststatistik:

$$T_n = (n-1)\frac{S_X^2}{\sigma_0^2}.$$

i) Wie ist die Prüfgröße unter H_0 verteilt? Wie lautet der bzw. lauten die Parameter der Verteilung?

ii) Geben Sie den kritischen Bereich mit konkreten Werten an.

iii) Wie lautet die Testentscheidung?

f) Bei einer Befragung gehen 70% der Personen davon aus, dass die relative Veränderung des Bruttoinlandsprodukts zum Vorjahr 2018 größer ist als 2016, 20% glauben sie wird den selben Wert annehmen, 10% vermuten einen Rückgang. Berechnen Sie die Gini-Entropie für diese Einschätzung.

2. In einem Onlineshop ist die Zeit t_i zwischen dem ersten Besuch eines Kunden i auf dem Internetauftritt und dem Kauf eines Produkts [in Minuten] gegeben:

i	1	2	3	4	5
t_i	13	7	29	5	11

Nehmen Sie an, dass diese Zeitdauern unabhängig und identisch verteilte Realisationen einer Exponentialverteilung mit Parameter λ sind. Die Log-Likelihood Funktion für dieses Modell ist damit gegeben durch:

$$LL(\lambda|t_1, \ldots, t_n) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} t_i$$

a) Wie lautet die Ableitung der Log-Likelihood Funktion bzgl. λ?

b) Zeigen Sie, dass der Maximum Likelihood Schätzer für λ durch $\widehat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} t_i}$ gegeben ist und berechnen Sie den konkreten Schätzwert für die Beobachtungen t_i. Hinweis: Gehen Sie davon aus, dass die hinreichenden Bedingungen für ein Maximum erfüllt sind.

Nehmen Sie im Folgenden an, dass den beschriebenen Wartezeiten bis zum Kauf eines Produkts tatsächlich eine Exponentialverteilung zu Grunde liegt und der Parameter $\lambda = 0.4$ ist. Die Dichtefunktion dieser Zufallsvariable X lautet $f(x) = \lambda \exp(-\lambda x)$, die Verteilungsfunktion $F(x) = 1 - \exp(-\lambda x)$.

c) Wie groß ist die Wahrscheinlichkeit, dass ein Kunde höchstens 4 Minuten nach dem erstem Besuch des Onlineshops einen Kauf tätigt?

d) Für Erwartungswert und Varianz von X gilt: $E(X) = \frac{1}{\lambda}$ und $Var(X) = \frac{1}{\lambda^2}$. Wie lautet das zweite Moment $E(X^2)$?

e) Es seien zwei exponentialverteilte Zufallsvariablen $X \sim Exp(\lambda = 3)$ und $Y \sim Exp(\lambda = 5)$ gegeben. Die Kovarianz zwischen X und Y betrage $Cov(X, Y) = 0.02$. Wie groß ist die Varianz von $X + Y$?

f) Berechnen Sie die Korrelation von X und Y.

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

In Ihrem Workspace liegt ein bereits erstellter Dataframe `data.sp500`, welcher den täglichen Kursverlauf des S&P 500 im Jahr 2016 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:

```
'data.sp500':        255 obs. of  5 variables:
 $ Close :  num  2012.66 2016.71 1990.26 1943.09 1922.03 1923.67 ...
 $ Open  :  num  2038.20 2013.78 2011.71 1985.32 1945.97 1926.12 ...
 $ Low   :  num  1989.68 2004.17 1979.05 1938.83 1918.46 1901.10 ...
 $ High  :  num  2038.20 2021.94 2011.71 1985.32 1960.40 1935.65 ...
 $ Volume:  num  4304880000 3706620000 4336660000 5076590000 ...
```

1. Geben Sie den `R`-Befehl an, mit dem Sie den arithmetrischen Mittelwert der Eröffnungskurse erhalten.

2. Geben Sie den `R`-Code an, mit dem Sie alle Tage mit einem Schlusskurs von mindestens 1500 und höchstens 3000 erhalten.

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

Folgender R-Code wird nun in die Konsole eingegeben:

```
> a <- data.sp500$High
> b <- mean(a[a>2500])
> c <- data.sp500$High-mean(data.sp500$High)
> d <- mean(c^2)
> e <- mean(c^4/sqrt(d)^4)
```

3. Welche statistischen Kennzahlen beschreiben die Variablen b und e in den obigen Befehlen?

Im Folgenden steht Ihnen in R der Data Frame `data.dax30` zur Verfügung, welcher die täglichen Kurse des Jahres 2016 der Aktien Allianz SE (`all`), BASF SE (`bas`), BMW AG (`bmw`), RWE AG (`rwe`) und Siemens AG (`sie`) enthält.[2] Arbeiten Sie zunächst mit folgendem R-Output:

```
'data.dax30':        260 obs. of  5 variables:
 $ all:  num  140.66 141.11 141.56 138.63 136.87 137.41 ...
 $ bas:  num  64.86 64.50 63.42 61.88 61.12 60.71 61.73 ...
 $ bmw:  num  88.35 87.94 85.03 81.83 79.91 79.63 81.26 ...
 $ rwe:  num  10.45 10.67 10.72 10.43 10.22 10.51 10.59 ...
 $ sie:  num  82.23 82.74 82.42 80.84 80.28 79.76 80.23 ...
```

4. Sie wissen, dass `cor(data.dax30$bmw,data.dax30$sie) = 0.5`, `var(data.dax30$bmw) = 1`, `var(data.dax30$sie) = 2` gilt. Berechnen Sie mit diesen Angaben die Stichprobenkovarianz zwischen den täglichen Kursen der Siemens AG und der BMW AG für das Jahr 2016.

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

5. Geben Sie den R-Code an, mit dem Sie die relative Kursentwicklung von dem ersten auf den vierten Tag für die BASF Aktie bestimmen können.

6. Mit welchem R-Code können Sie sich für die Siemens AG Aktie die Spannweite berechnen lassen?

7. Vervollständigen Sie die folgende R-Funktion so, dass damit aus einer Stichprobe `stich` mit beliebigem Stichprobenumfang n ein linksseitiges exaktes Konfidenzintervall für den Parameter λ einer exponentialverteilten Grundgesamtheit zum Niveau `alpha` berechnet werden kann.

```
Exp_exact_KI=function(         ,alpha){
Xn = mean(stich)
n =          (stich)
lower = 0
upper =          chisq(p =         , df = 2*n) / (2*n*Xn)
KI=c(lower,upper)
return(KI)}
```

Folgender R-Code wird nun in die Konsole eingegeben:

```
> j = 1
> for(i in 0:3){j = j * i}
```

8. Welchen Wert erhalten Sie für j?

9. Gegeben sei $Y \sim Bin\,(100, 0.5)$. Mit welchem `R`-Code können Sie $P\,(20 < Y \leq 70)$ berechnen?

In ihrem Workspace befindet sich folgende Funktion:

```
myfun=function(g1,g2,e1,e2,e3){
  g1*e1+g2*e2+(1-g1-g2)*e3
}
```

10. Nehmen Sie an, $e1$, $e2$ und $e3$ seien Mittelwerte und weiterhin gelte $g1, g2 \in (0, 1)$ mit $g1 + g2 < 1$. Was wird mit Hilfe der Funktion `myfun` berechnet?

11. Vervollständigen Sie den nachfolgenden `R`-Code zur Optimierung der Funktion `myfun` hinsichtlich g2 mit nlm.

```
nlm(          ,0.05, e1 = 2, e2 = 4, e3 = 3.5, g1 = 0.3)
```

Schmierpapier zu Aufgabe 4

Lösung Klausur WiSe 17/18 (10 ECTS)

Lösung 1

1. $T_n = \sqrt{n}\frac{\bar{Z}_n - \mu_0}{\sigma}$ $\quad E[T_n] = 0, Var[T_n] = 1$ 1P
2. $|t_n| > \lambda_{1-\alpha/2}$ $\quad \lambda_{0.995} = 2.5758$ 1P
3. 50 0.5P
4. 7.0711 0.5P
5. $t_n > \lambda_{1-\alpha/2} = 2.5758$
   H_0 kann bei Irrtums-WS von $\alpha = 1\%$ abgelehnt werden. 1P
6. $p < \alpha$, da H_0 abgelehnt wird. 0.5P
7. $\pi = 2 - \Phi(1.87) - \Phi(3.28) = 0.0312$ 1P
8. Erhöhung von Stichprobenumfang n 0.5P
9. 0 0.5P
10. $N(80.000, 10.000)$ 1P
11. $\Phi(\frac{38.000 - 38.300}{\sqrt{12.000}}) = \Phi(-2.74) = 1 - \Phi(2.74) = 0.0031$ 1.5P
12. 38.300 0.5P
13. 1 0.5P

Lösung 2

1a) 0.09 — 0.5 P

1b) $F_Y(1) = 0.33$ und $F_Y(2.5) = 0.68$ — 1P

1c) $P(Y \geq 2) = 1 - P(Y \leq 1) = 0.67$ — 1P

2a) $\mathbb{E}[X] = \lambda \Leftrightarrow \hat{\lambda}_{MM} = M_1' = \frac{1}{n} \sum_{i=1}^{n} X_i^1 = \bar{X}_n$ und $\hat{\lambda} = 4$ — 1P

2b) $E[\bar{X}_n] = E[X] = \lambda$ — 1P

3. $E[R_n] = \rho$ — 0.5 P

4a) 15 — 0.5 P

4b) $\sum_{i=1}^{5} X_i = 20 > 15 \Rightarrow$ Lehne H_0 bei einer Irrtumswahrscheinlichkeit von 4.87% ab. 1 P

4c) 25% — 0.5P

5a) $P(3 \leq X \leq 5) = 0.7851 - 0.2381 = 0.547/0.5471$ — 1P

5b) $Var(X) = 4$ — 0.5P

5c) $Y \sim Pois(8)$ — 0.5P

5d) $Z \sim Pois(8)$ — 0.5P

6. 5 — 0.5P

Lösung 3

1. a) Median: 0.03 — 0.5 P

   b) bleibt gleich — 0.5 P

   c) Geometrisches Mittel: 1.0252 — 0.5 P

   d) Stichprobenvarianz: 0.0008 — 1 P

   e) i.) Prüfgröße ist χ^2 verteilt mit Parameter $n-1=8$. — 0.5 P

   e) ii.) Lehne H_0 ab, wenn $t_n < \chi^2_{0.05;8} = 2.73$. — 1 P

   e) iii.) Lehne H_0 ab, wenn $t_n < \chi^2_{0.05;8} \Rightarrow 8 \cdot \frac{0.0006}{0.001} = 4.8 > 2.73$
   $\Rightarrow H_0$ beibehalten — 1 P

   f) Gini-Entropie: $0.7 \cdot 0.3 + 0.2 \cdot 0.8 + 0.1 \cdot 0.9 = 0.46$ — 0.5 P

2. a) $\frac{\partial LL(\lambda|t_1,\ldots,t_n)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i$ — 1 P

   b) $\frac{\partial LL(\lambda|t_1,\ldots,t_n)}{\partial \lambda} = 0 \Leftrightarrow \widehat{\lambda}_{\text{ML}} = \frac{n}{\sum_{i=1}^{n} t_i}, \quad \widehat{\lambda}_{\text{ML}} = 0.0769$ — 1 P

   c) $P(X \leq 4) = 1 - \exp(-0.4 \cdot 4) = 0.7981$ — 0.5 P

   d) $E(X^2) = Var(X) + E(X)^2 = \frac{2}{\lambda^2}$ — 1 P

   e) $Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y) = 0.1911$ — 0.5 P

   f) $Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = 0.3$ — 0.5 P

Lösung 4

1. `mean(data.sp500$Open)` 0.5 P
2. `which(data.sp500$Close >=1500 & data.sp500$Close <=3000)` 1.0 P
3. bedingter Mittelwert, Kurtosis 0.5 + 0.5 P
4. 0.7071 1.0 P
5. `data.dax30$bas[4]/data.dax30$bas[1]-1` 1.0 P
6. `max(data.dax30$sie)-min(data.dax30$sie)` 0.5 P
7. stich, length, q, 1-alpha 0.5+0.5+0.5+1 P
8. 0 0.5 P
9. pbinom(70,100,0.5)-pbinom(20,100,0.5) 1 P
10. gewichteter Mittelwert 0.5 P
11. myfun 0.5 P