Klausur Statistik (10 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	13.08.2019
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

Eine Maschine stellt Schokoriegel her. Das Gewicht in Gramm G_1 der einzelnen Riegel ist dabei unabhängig und identisch normalverteilt mit Mittelwert 55 und Varianz 0.5.

1. Wie groß ist die Wahrscheinlichkeit, dass ein produzierter Schokoriegel weniger als 54.9 g wiegt?

2. Wie groß ist die Wahrscheinlichkeit, dass ein produzierter Schokoriegel zwischen 54.8 g und 55.2 g wiegt?

Unabhängig davon produziert eine zweite Maschine Schokowaffeln. Das Gewicht in Gramm G_2 dieser Waffeln ist ebenfalls unabhängig und identisch normalverteilt, allerdings mit Mittelwert 30 und Varianz 0.75.

3. Wie ist das Gewicht einer Packung $G_P = 5G_1 + 5G_2$ von jeweils 5 Schokoriegeln und -waffeln verteilt? Geben Sie die Verteilungsfamilie, sowie Erwartungswert und Varianz explizit an.

Im Folgenden sind die absoluten Häufigkeiten für die Bewertungen eines Hotels im Jahr 2017 angegeben. Dabei wird die Skala 1 (sehr gut), 2 (gut), 3 (mittel) und 4 (schlecht) zu Grunde gelegt.

Bewertung des Hotels	1	2	3	4
Absolute Häufigkeit	5	3	2	1

4. Geben Sie das Skalenniveau des Merkmals "Bewertung des Hotels" an.

5. Geben Sie den Modus des Merkmals "Bewertung des Hotels" an.

6. Berechnen Sie die Gini-Entropie des Merkmals "Bewertung des Hotels" aus den relativen Häufigkeiten.

7. Geben Sie Minimal- und Maximalwert an, den die Gini-Entropie aus Teilaufgabe 6. annehmen könnte.

Die Zufallsvariable X besitzt den Erwartungswert $E(X) = 3$ und die Varianz $Var(X) = 4$.

8. Bestimmen Sie das zweite Moment $E(X^2)$.

9. Der erste Pearsonsche Schiefekoeffizient von X beträgt 0.5. Bestimmen Sie den Modus von X.

Eine zweite Zufallsvariable Y besitzt die Varianz $Var(Y) = 9$ und die Korrelation zwischen X und Y beträgt $Cor(X, Y) = 0.5$.

10. Bestimmen Sie die Kovarianz $Cov(X, Y)$.

11. Welchen Wert müsste $Cov(X, Y)$ annehmen, damit die Korrelation ihren Maximalwert erreicht?

12. Nehmen Sie nun an, dass die Kovarianz $Cov(X, Y) = 2$ beträgt und $E(X \cdot Y) = 1$ ist. Wie lautet der Erwartungswert $E(Y)$?

Die Zufallsvariable Z ist exponentialverteilt mit Parameter $\lambda = 5$ und besitzt die Verteilungsfunktion $F(z) = 1 - \exp(-5z)$.

13. 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 1

Aufgabe 2 von 4

Als Geschäftsführer der Einzelhandelskette *Nordi* interessieren Sie sich für den Krankenstand Ihrer Mitarbeiter. Sie gehen davon aus, dass die Wartezeit X in Tagen bis zur Erkrankung eines Mitarbeiters exponentialverteilt mit Parameter $\lambda = 0.2$ ist.

1. Bestimmen Sie den Wert der Verteilungsfunktion $\mathrm{F}_X(x)$ an der Stelle $x = 2.5$.

2. Wie groß ist die Wahrscheinlichkeit, dass die Wartezeit bis zum Erkranken eines Mitarbeiters mehr als 7 Tage beträgt?

3. Wie groß ist die Wahrscheinlichkeit, dass zwei Mitarbeiter gleichzeitig erkranken?

4. Wie groß ist die Wahrscheinlichkeit, dass die Wartezeit bis zum Erkranken eines Mitarbeiters mindestens 1 und höchstens 3 Tage beträgt?

5. Überprüfen Sie, ob die Dichtefunktion der Zufallsvariable X symmetrisch um den Wert 0.2 ist.

6. Berechnen Sie die Varianz von X für $\lambda = 0.2$.

7. Die Wahrscheinlichkeit, dass das Stichprobenmittel $\overline{X}_n$ um höchtens 1.2 vom wahren Erwartungswert abweicht, soll mindestens 97.9% betragen. Bestimmen Sie mit der Ungleichung von Chebyshev den minimalen Umfang n einer Zufallsstichprobe $X_1, ..., X_n$, der hierfür benötigt wird.
   Hinweis: Falls Sie Teilaufgabe 6. nicht lösen konnten, verwenden Sie $Var[X] = 25$.

Sie erfassen 10 mal die Wartezeit bis zur Erkrankung eines Mitarbeiters. Es ergibt sich $\sum_{i=1}^{10} x_i = 72$ und $\sum_{i=1}^{10}(x_i - \overline{x}_{10})^2 = 68$.

8. Bestimmen Sie die Stichprobenvarianz.

9. Schätzen Sie die tatsächliche durchschnittliche Wartezeit μ_X geeignet.

10. Ihr Kollege zweifelt an, dass die Wartezeit X der Exponentialverteilung mit Parameter $\lambda = 0.2$ folgt. Vielmehr vermutet er, dass $\lambda > 0.3$ gilt. Er stellt das Hypothesenpaar

    $$H_0 : \lambda \leq 0.3 = \lambda_0 \ \text{ vs. } \ H_1 : \lambda > 0.3$$

    auf. Treffen Sie mittels der exakten Entscheidungsregel eine Testentscheidung (mit Begründung!) für eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$.
    Erinnerung: $\sum_{i=1}^{10} x_i = 72$ und $\sum_{i=1}^{10}(x_i - \overline{x}_{10})^2 = 68$.

11. Würden Sie für den von Ihnen durchgeführten Test einen p-Wert von 0.004, 0.04, oder 0.4 erwarten? Begründen Sie Ihre Antwort.

11. Gehen Sie nun davon aus, dass X näherungsweise normalverteilt ist, wobei die Varianz unbekannt ist. Berechnen Sie ein zweiseitiges 90%-Konfidenzintervall für die tatsächliche durchschnittliche Wartezeit.
Verwenden Sie: $s_{10}^2 = 6.8$ und $\overline{x}_{10} = 7.2$.

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

Gegeben sei die Zufallsvariable X: "Anzahl der Personen, die in Nürnberg innerhalb einer Stunde aufgrund überhöhter Geschwindigkeit geblitzt werden" mit $X \sim Pois(5)$. Gehen Sie davon aus, dass die Annahmen eines Poisson-Prozesses erfüllt sind.

1. Berechnen Sie die Wahrscheinlichkeit, dass innerhalb einer Stunde kein Auto geblitzt wird.

2. Berechnen Sie die Wahrscheinlichkeit, dass innerhalb einer Stunde mehr als 3 aber höchstens 7 Personen geblitzt werden.

3. Berechnen Sie die Wahrscheinlichkeit, dass in der ersten Stunde 3 und in der zweiten Stunde 6 Personen geblitzt werden.

4. Geben Sie den Median der theoretischen Verteilung von X an.

5. Geben Sie das 99.99% Quantil der theoretischen Verteilung von X an.

6. Welcher Verteilung folgt die Zufallsvariable Z : “Anzahl der Personen, die in Nürnberg innerhalb von zwei Stunden aufgrund überhöhter Geschwindigkeit geblitzt werden”.

7. Welcher allgemeine Zusammenhang besteht zwischen dem Erwartungswert und der Varianz einer Poissonverteilung.

Bei einer Verkehrszählung wird die Anzahl der Autos Y, die innerhalb einer Stunde durch die Lange Gasse fahren, gezählt. Sie machen die folgenden 80 Beobachtungen:

i	1	2	3	4	5	6
y_i	0	1	2	3	4	≥ 5
n_i	16	25	23	6	10	0

n_i bezeichne die Anzahl der Beobachtungen für das Ereignis y_i.

8. Berechnen Sie das Stichprobenmittel $\bar{Y}_n$.

Verwenden Sie ab jetzt $\bar{Y}_n = 1.5$.

9. Geben Sie unter der Annahme, dass Y poissonverteilt ist, das approximative, symmetrische 95% Konfidenzintervall für den Parameter λ an und berechnen Sie die Grenzen.

10. Schätzen Sie für $y_2 = 1$ die Einzelwahrscheinlichkeit $\hat{p}_1 = f_{Pois}(y_2; \hat{\lambda})$. (Hinweis: $\hat{\lambda} = \bar{Y}_n$)

Sie wollen nun Ihre Annahme einer Poissonverteilung mit unbekanntem Parameter mittels eines χ^2-Anpassungstests überprüfen. Hierzu haben Sie folgende Werte gegeben (Hinweis: Klassen sind bereits zusammengeführt, um die Approximationseigenschaften zu erfüllen):

i	1	2	3	4	5
y_i	0	1	2	3	≥ 4
$\hat{p}_i^0$	0.2231	0.3347	0.2510	0.1255	0.0657
$\frac{(N_i - n\hat{p}_i^0)^2}{n\hat{p}_i^0}$	0.1913	0.1178	0.4246	1.6257	4.2819

11. Geben Sie die theoretische Teststatistik inklusive asymptotischer Verteilung unter der Nullhypothese an und berechnen Sie die realisierte Teststatistik.

12. Geben Sie die kritische Schranke des Tests für eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$ an.

13. Treffen Sie eine Testentscheidung (mit Begründung). Falls Sie die Teilaufgaben 11. und 12. nicht lösen konnten, verwenden Sie für die Teststatistik den Wert 7 und für die kritische Schranke den Wert 8.

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

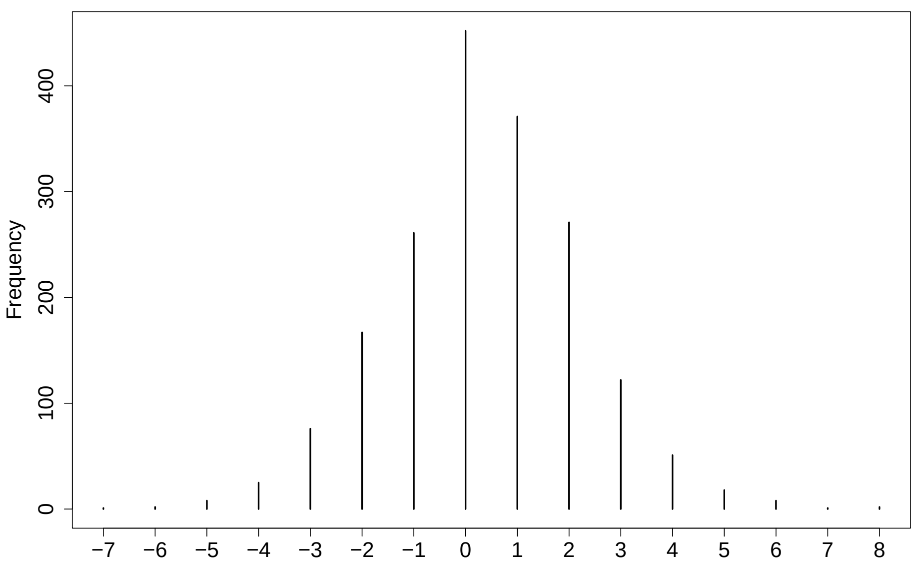
In Ihrem Workspace liegt ein Dataframe `data`, welcher folgende Informationen zu 1836 Fußballspielen enthält:

Spaltenname	**Info**
`Season`	Saison
`Home`	Heimmannschaft
`Away`	Auswärtsmannschaft
`Goal_h`	Anzahl der Tore der Heimmannschaft
`Goal_a`	Anzahl der Tore der Auswärtsmannschaft

Der Output und die Grafik werden durch die folgenden Befehle erstellt:

```
data[c(1:6,1800),]
variable=data$Goal_h-data$Goal_a
plot(table(variable))
```

```
        Season                 Home                 Away     Goal_h     Goal_a
1    2012-2013    Borussia Dortmund        Werder Bremen          2          1
2    2012-2013 Eintracht Frankfurt     Bayer Leverkusen          2          1
3    2012-2013           FC Augsburg Fortuna Duesseldorf          0          2
4    2012-2013          Hamburger SV      1. FC Nuernberg          0          1
5    2012-2013           Hannover 96        FC Schalke 04          2          2
6    2012-2013     Moenchengladbach     1899 Hoffenheim          2          1
1800 2017-2018        Werder Bremen           RB Leipzig          1          1
```



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

1. Geben Sie den `R`-Code für die Berechnung der maximalen Anzahl an Toren der Auswärtsmannschaft an.

2. Geben Sie den `R`-Code für die Berechnung der absoluten Häufigkeiten der Anzahl der Tore der Heimmannschaft an.

3. Nehmen Sie Stellung zu folgender Aussage (mit Begründung!): Der höchste beobachtete Heimsieg lautet 12:1.

4. Geben Sie den `R`-Code an, mit dem Sie alle Daten aus der Saison 2012-2013 erhalten.

5. Geben Sie den `R`-Code an, mit dem Sie die Anzahl der Spiele erhalten, bei denen die Auswärtsmannschaft mehr als zwei Tore schießt.

6. Welchen Output liefert der Befehl `data$Home[c(rep(2,1),seq(3,5,by=2))]`?

7. Mit welchem `R`-Code können Sie den Korrelationskoeffizienten zwischen der Anzahl der Tore der Heimmannschaft und der Auswärtsmannschaft bestimmen?

8. Vervollständigen Sie den folgenden `R`-Code, um eine Funktion zu erhalten, die den Mittelwert von x berechnet bedingt darauf, dass x größer als a und kleiner als b ist. Es gilt $b > a$ mit $a, b \in \mathbb{R}$.

```
f3 = function(x,a,b){
   c <- ________(x[x ________ ________ ________ x ________ ________])
   return(c)
}
```

9. Geben Sie den R-Code an, mit dem Sie die Anzahl der Spiele erhalten, bei denen entweder die Heimannschaft, die Auswärtmannschaft oder beide kein Tor erzielt hat.

Folgender `R`-Code wird nun in die Konsole eingegeben:

```
> j = 2
> for(i in 1:3){ if(i==j){j = j + i}}
```

10. Welchen Wert erhalten Sie für j?

11. Gegeben sei $Y \sim Bin\,(100, 0.6)$. Mit welchem `R`-Code können Sie $P\,(40 < Y \leq 80)$ berechnen?

12. Welcher der folgenden `R`-Codes berechnet die logarithmierte Likelihoodfunktion der Poissonverteilung mit Parameter `lambda`. Tragen Sie den entsprechenden Buchstaben in das Feld ein.

```
A: LL_poi=function(lambda,data){log(sum(dpois(lambda,data)))}
B: LL_poi=function(lambda,data){log(dpois(sum(data,lambda)))}
C: LL_poi=function(lambda,data){dpois(log(sum(lambda,data)))}
D: LL_poi=function(lambda,data){sum(log(dpois(data,lambda)))}
```

Schmierpapier zu Aufgabe 4