# Klausur Statistik (7.5 ECTS)

| | | | |
|---|---|---|---|
| **Name** | | **Prüfer** | Prof. Dr. I. Klein |
| **Vorname** | | **Arbeitszeit** | Freitag, 19.02.2016 |
| **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 seit dem WS 2014/15 offiziell herausgegebene Formelsammlung (DIN A5, gebunden, orangener Umschlag). Es sind prinzipiell keine weiteren Eintragungen oder Markierungen darin erlaubt. Ausgenommen sind farbliche Hinterlegungen von Textpassagen und/oder Formeln. Außerdem ist das Erratum zur 1. Auflage der Formelsammlung 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

Anlässlich der Fußball-Europameisterschaft fahren die begeisterten Fußball-Fans Lukas und Basti mit dem Fanbus nach Frankreich. Sie vergleichen die beiden Länder Deutschland und Frankreich. In angegebener Tabelle sind ihre Ergebnisse festgehalten.[1]

| | Deutschland | Frankreich |
|---|---|---|
| Hauptstadt | Berlin | Paris |
| Bevölkerung in Tausend | 81174 | 66352 |
| Fläche in $km^2$ | 357137 | 632834 |
| Bruttoinlandsprodukt (BIP) je Einwohner in EUR | 36000 | 32200 |

1. Nennen Sie den höchsten Skalentyp des Merkmals 'Hauptstadt'.

2. Die Bevölkerung in Deutschland steigt im nächsten Jahr um 2% und sinkt im übernächsten Jahr um 2%. Wie groß ist die Bevölkerungszahl in zwei Jahren?

3. Wie hoch ist das Bruttoinlandsprodukt je Einwohner für beide Länder zusammen?

4. Der Bus fährt den ersten Streckenabschnitt, welcher eine Länge von 350 km besitzt, mit einer Durchschnittsgeschwindigkeit von 100 km/h. Der zweite Abschnitt beträgt 400 km und wird mit einer Durchschnittsgeschwindigkeit von 80 km/h zurückgelegt. Berechnen Sie die durchschnittliche Geschwindigkeit auf der gesamten Strecke.

[1] https://www.destatis.de/Europa/DE/Staat/EUStaaten/Deutschland.html , https://www.destatis.de/Europa/DE/Staat/EUStaaten/Frankreich.html

Schiedsrichter Walter Streng zieht im Mittel 4.5 Karten während eines Fußballspiels, wobei nicht zwischen gelben und roten Karten unterschieden werden soll. Die Anzahl der pro Spiel gezogenen Karten soll durch die Poisson-verteilte Zufallsvariable $Y$ beschrieben werden. Im Folgenden werden nur Fußballspiele mit einer Dauer von 90 Minuten betrachtet, das heißt es gibt keine Verlängerung, kein Elfmeterschießen oder Ähnliches.

5. Welchen Wert nimmt der Parameter für die Verteilung von $Y$ an?

**Arbeiten Sie jetzt mit $\underline{\boldsymbol{\lambda = 3}}$ weiter!**

6. Mit welcher Wahrscheinlichkeit zeigt der Schiedsrichter in einem Spiel genau 2 Karten?

7. Berechnen Sie die Wahrscheinlichkeit, dass der Schiedsrichter im ersten Spiel mehr als 2 Karten und im zweiten Spiel höchstens 3 Karten vergibt. Gehen Sie hierbei davon aus, dass die Anzahl der Karten im ersten Spiel ($Y_1$) unabhängig von der Anzahl der Karten im zweiten Spiel ($Y_2$) ist.

8. Die Zufallsvariable $W$ beschreibt die Wartezeit zwischen zwei Ereignissen in Minuten. Nationalspieler Thomas Nüller hat gerade die gelbe Karte gesehen. Berechnen Sie die Wahrscheinlichkeit, dass innerhalb der nächsten 10 Minuten ein beliebiger Spieler die nächste Karte sieht.

Der Trainer hat die Hypothese, dass all seine 23 Spieler eine Schussgeschwindigkeit haben, deren Varianz nicht größer als 64 ist. Sie testen die Hypothese mit einem geeigneten Hypothesentest, wobei die Irrtumswahrscheinlichkeit 5% beträgt. Dafür betrachten Sie die Schussgeschwindigkeiten der 23 Spieler $X_1, X_2, \ldots, X_{23}$ als stochastisch unabhängig, identisch normalverteilt. Sie stellen folgende Hypothesen auf:

$$H_0 : \sigma^2 \leq \sigma_0^2 = 64 \qquad \text{gegen} \qquad H_1 : \sigma^2 > \sigma_0^2 = 64.$$

9. Zeigen Sie, dass die folgende Gleichung gilt: $\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2$.

10. Bestimmen Sie die realisierte Prüfgröße, wenn für die Stichprobe $\sum_{i=1}^{23} x_i^2 = 309570$ und $\overline{x}_{23} = 116$ gelten.

**Verwenden Sie im Folgenden als Wert für die realisierte Prüfgröße 5 und als Wert für die kritische Schranke 40!**

11. Treffen Sie die Testentscheidung und begründen Sie diese.

**Schmierpapier zu Aufgabe 1**

# Aufgabe 2

Nach der Überschwemmung der Stadt P vor zwei Jahren versuchen Sie den Stadtrat vom Erwerb eines mobilen Hochwasserschutzes zu überzeugen. Sei

$H$ : „Höhe des Hochwassers, falls der Fluss D sein Flussbett verlässt" [m].

Aus einer einfachen Stichprobe $H_1, ..., H_{100}$ aus den letzten 100 Jahre haben Sie sich bereits folgende Daten extrahiert:

$$\sum_{i=1}^{100} h_i = 102.30 \qquad \text{und} \qquad \sum_{i=1}^{100} h_i^2 = 190.43.$$

1. Was bedeutet „erwartungstreu"?

2. Geben Sie einen erwartungstreuen Schätzer für die Varianz an und berechnen Sie den zugehörigen Schätzwert.

3. Nennen Sie eine Eigenschaft eines Momenten-Schätzers.

Sie wissen, dass

$$P(H \leq x) = 1 - \exp(-x) \quad \text{mit zugehöriger Dichtefunktion} \quad f(x) = \exp(-x)$$

für $x \geq 0$.

4. Der Stadtrat behauptet, dass die Wahrscheinlichkeit, dass das Hochwasser die gängigen Barrikaden überschreitet, sehr gering sei. Berechnen Sie

   a) die Wahrscheinlichkeit, dass der Hochwasserstand genau 3 m beträgt,

   b) die Wahrscheinlichkeit, dass der Hochwasserstand eine Barrikade von 1 m überschreitet,

   c) die Wahrscheinlichkeit, dass der Hochwasserstand maximal 4 m beträgt, wenn er die 3 m Marke bereits überschritten hat.

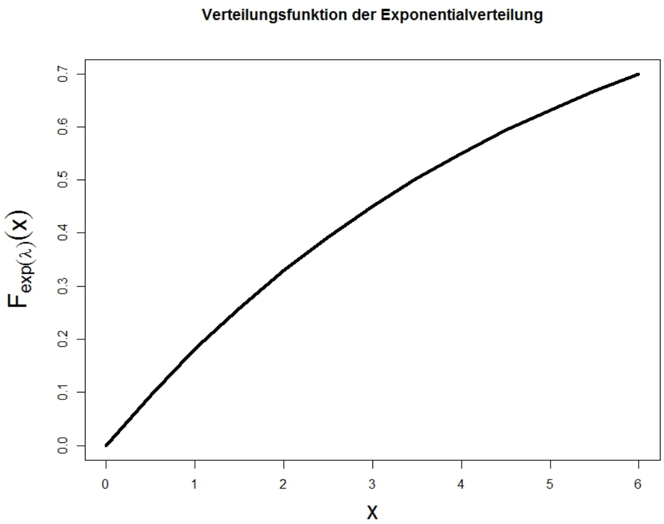
Sie drängen den Stadtrat zu einer schnellen Entscheidung, da ihrer Meinung nach der Fluss D im Mittel alle 5 Jahre eine Überschwemmung verursacht. Sie wissen, dass

$Z$ : „Zeit zwischen zwei Überschwemmungen in Jahren"

exponentialverteilt ist mit Parameter $\lambda$.

5. Geben Sie auf Grundlage der erhaltenen Informationen den Parameter $\lambda$ an.

6. Markieren Sie den Median der Exponentialverteilung in der nachstehenden Graphik eindeutig.




Ein Mitglied des Stadtrates widerspricht Ihnen und äußert die Vermutung, dass der Fluss D seltener über die Ufer tritt. Sie betrachten hierzu eine einfache Stichprobe $Z_1, ..., Z_{20}$ mit $\sum_{i=1}^{20} z_i = 51.75$.

7. a) Geben Sie das theoretische und das realisierte **rechtsseitige** Konfidenzintervall für den Parameter $\lambda$ zum Niveau $1 - \alpha = 0.95$ an.

b) Interpretieren Sie das realisierte Konfidenzintervall.

8. Testen Sie nun auf dem 95%-Niveau Ihre Vermutung gegen die Ihres Kontrahenten in Form der Hypothesen

$$H_0: \lambda \leq \frac{1}{5} = \lambda_0 \qquad \text{gegen} \qquad H_1: \lambda > \frac{1}{5}$$

mit zugehöriger Prüfgröße

$$T_n = 2\lambda_0 \sum_{i=1}^{n} Z_i \sim \chi^2_{2n}$$

a) Berechnen Sie die realisierte Prüfgröße $t_{20}$.

b) Berechnen Sie den p-Wert. **Sollten Sie in Aufgabe 8a) keinen Wert erhalten haben, verwenden Sie $t_{20} = 22.16$.**

c) Treffen Sie eine Testentscheidung und begründen Sie diese in Form einer Entscheidungsregel.

**Schmierpapier zu Aufgabe 2**

# Aufgabe 3

Ihr Freund Kevin ist begeisterter Sportler. Sie besuchen Kevin bei seinem Training.

1. Zuerst beobachten Sie den Weitsprung. Nehmen Sie zunächst an, dass beim Weitsprung-Training im Allgemeinen die Ergebnisse der gesprungenen Weiten (Zufallsvariable W) in Meter unabhängig und identisch normalverteilt sind mit $W \sim N(\mu = 6.37, \sigma^2 = 0.04)$.

   a) Berechnen Sie die Grenzen des zweifachen zentralen Schwankungsintervalls für die Sprungweite W.

   b) Mit welcher Wahrscheinlichkeit liegt die gesprungene Weite eines Sportlers im 1.53-fachen zentralen Schwankungsintervall?

   c) Mit welcher Wahrscheinlichkeit springt ihr Freund Kevin mindestens eine Weite von 6.50 Metern?

   Nun notieren Sie sich die gesprungenen Weiten (W) in Meter der ersten 10 Sprünge von Kevin:

   | **Sprung i** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
   |---|---|---|---|---|---|---|---|---|---|---|
   | $w_i$ | 6.30 | 6.29 | 6.51 | 6.22 | 6.49 | 6.27 | 6.46 | 6.18 | 6.40 | 6.36 |

   d) Berechnen Sie die durchschnittliche gesprungene Weite der 10 beobachteten Sprünge.

e) Berechnen Sie die Spannweite der beobachteten gesprungenen Weiten.

2. Als nächstes wenden Sie sich dem Training im Sprint zu und beobachten 20 Sprinter. Nehmen Sie an, die „Zeiten im 100m-Sprint [in Sekunden]" (X) sind normalverteilt mit bekannter Varianz $\sigma^2$.

   Es wird der Maximum-Likelihood-Schätzer für den unbekannten Parameter $\mu$ gesucht.

   a) Zeigen Sie, dass die logarithmierte Likelihoodfunktion von $\mu$ gegeben ist durch

   $$\ln(L(\mu; X_1, ..., X_n)) = -n \ \ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2}.$$

   b) Zeigen Sie, dass der Maximum-Likelihood-Schätzer für den unbekannten Parameter $\mu$ lautet: $\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

   c) Berechnen Sie den Wert von $\hat{\mu}_{ML}$, wenn Sie wissen, dass $\sum_{i=1}^{20} x_i = 179$.

d) Formulieren Sie in wenigen Worten die hinreichende Bedingung für das Vorliegen eines globalen Maximums an der Stelle $\hat{\mu}_{ML}$.

Zum Abschluss seines heutigen Trainings macht Kevin noch ein paar Klimmzüge. Sie beobachten ihn und stellen fest, dass die Klimmzugstange schon sehr alt ist.

3. Nehmen Sie an, die Wahrscheinlichkeit, dass die Klimmzugstange zerbricht, sei bei jedem einzelnen Klimmzug gleichbleibend $p = \frac{1}{43}$. Außerdem besitze die Anzahl der Klimmzüge (Z) bevor die Klimmzugstange zerbricht folgende Wahrscheinlichkeitsfunktion:

$$f_Z(z) = p(1-p)^z.$$

a) Wie hoch ist die Wahrscheinlichkeit, dass die Stange bei Kevins erstem Klimmzug zerbricht?

b) Wie nennt sich die Verteilung von Z?

c) Wie viele Klimmzüge sind zu erwarten, bevor die Klimmzugstange zerbricht?

**Schmierpapier zu Aufgabe 3**

# Aufgabe 4

Die Präzisionswerke $W_1$ und $W_2$ stellen ein Bauteil für Maschinen her. Bei der Qualitätsprüfung werden die Bauteile in *fehlerhaft* ($F$) und *fehlerfrei* ($\overline{F}$) unterschieden.

1. Es wurden 20% der Bauteile von $W_1$ bezogen. Alle weiteren Bauteile wurden von $W_2$ bezogen. Von den Bauteilen, die von $W_1$ bezogen wurden, waren 3% fehlerhaft. Von den Bauteilen, die von $W_2$ bezogen wurden, waren 0.7% fehlerhaft.

   a) Stellen Sie den Sachzusammenhang in einer vollständig ausgefüllten 4-Feldertafel dar.

   b) Wie groß ist der Anteil der fehlerfreien Bauteile unter allen bezogenen Bauteilen?

   c) Wie groß ist der Anteil der fehlerfreien Bauteile unter den fehlerfreien Bauteilen aus Werk $W_2$?

d) Wie groß ist der Anteil der von Werk $W_2$ bezogenen Bauteile unter den fehlerhaften Bauteilen?

Das Präzisionswerk $W_3$ bietet ebenfalls das Bauteil an. Es soll entschieden werden, ob das Bauteil in Zukunft auch von $W_3$ bezogen wird. Dazu ist eine Probebestellung geplant. Mittels einer Stichprobe mit Zurücklegen soll auf den Anteil der fehlerhaften Bauteile von $W_3$ geschlossen werden.

Mit

$$X_i = \begin{cases} 1 & \text{, falls } i\text{-tes Bauteil fehlerhaft} \\ 0 & \text{, sonst} \end{cases}$$

ergibt sich die einfache Stichprobe $X_1, \ldots, X_n$, wobei $p$ die Wahrscheinlichkeit ist, ein fehlerhaftes Bauteil auszuwählen.

2. Welcher Zusammenhang besteht zwischen dem Wert des realisierten Stichprobenmittels $\overline{x}_n$ und dem Anteil der fehlerhaften Bauteile unter den ausgewählten Bauteilen?

3. Geben Sie die parametrisierte Verteilung von $\sum_{i=1}^{n} X_i$ an.

4. Bestimmen Sie mittels des zentralen Grenzwertsatzes den minimalen Stichprobenumfang für den Anteil der fehlerhaften Bauteile aus Werk $W_3$ so, dass die Wahrscheinlichkeit für eine Abweichung um höchstens 2 Prozentpunkte vom wahren Anteilswert mindestens 90% beträgt.

5. Bei der Bestimmung der minimalen Stichprobenlänge für Anteilswerte steht der unbekannte Anteilswert $p$ in der Varianz der Grundgesamtheit $p(1-p)$. Deshalb wird die Varianz in der Grundgesamtheit durch $p(1-p) \leq \frac{1}{4}$ abgeschätzt.

   a) Für welchen Wert $p$ wird diese Obergrenze angenommen?

   Gehen Sie nun davon aus, dass der wahre Anteil der fehlerhaften Bauteile kleiner oder gleich 0.1 ist.

   b) Zeigen Sie, dass mit dieser Zusatzinformation die Varianz in der Grundgesamtheit sogar durch $p(1-p) \leq \frac{9}{100}$ abgeschätzt werden kann.
   **Hinweis**: Zeigen Sie, dass $f(p) = p(1-p)$ im relevanten Bereich $p \in ]0, 0.1]$ streng monoton zunehmend ist und schließen damit auf den maximalen Wert.

   c) Bestimmen Sie den minimalen Stichprobenumfang für die Werte $\alpha = 0.1$ und $\varepsilon = 0.02$ mittels Chebyshev-Ungleichung. Passen Sie dazu die Formel so an, dass die Zusatzinformation $p \leq 0.1$ berücksichtigt wird.

**Schmierpapier zu Aufgabe 4**

# Lösung Klausur WS1516 (7.5 ECTS)

## Lösung 1

1) Nominalskala 0.5 P

2) $81174 \cdot 1.02 \cdot 0.98 = 81141.5304$ 0.5 P

3) $\dfrac{36000 \cdot 81174 + 32200 \cdot 66352}{81174 + 66352} = 34290.8938$ 1.0 P

4) $\dfrac{350 + 400}{\frac{350}{100} + \frac{400}{80}} = 88.2353$ 1.0 P

5) $\lambda = 4.5$ 0.5 P

6) $P(Y_1 = 2) = 0.2240$ 0.5 P

7) $P(Y_1 > 2)P(Y_2 \leq 3) = (1 - 0.4232) \cdot 0.6472 = 0.3733$ 1.5 P

8) $F_{exp}(10; \frac{3}{90}) = 1 - exp(-\frac{3}{90} \cdot 10) = 0.2835$ 1.5 P

9) $\sum\limits_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum\limits_{i=1}^{n} x_i^2 - \underbrace{\sum_{i=1}^{n} 2x_i}_{=2n\overline{x}_n} \overline{x}_n + \underbrace{\sum_{i=1}^{n} \overline{x}_n^2}_{=n\overline{x}_n^2} = \sum\limits_{i=1}^{n} x_i^2 - 2n\overline{x}_n^2 + n\overline{x}_n^2 = \sum\limits_{i=1}^{n} x_i^2 - n\overline{x}_n^2$ 1.0 P

10) $t_n = \dfrac{\sum\limits_{i=1}^{n}(x_i - \overline{x}_n)^2}{\sigma_0^2} = \dfrac{309570 - 23 \cdot 116^2}{64} = 1.2813$ 1.0 P

11) Lehne $H_0$ nicht ab, da $PG < KS$ (PG nicht im KB). 1.0 P

# Lösung 2

1. Der Erwartungswert des Schätzers entspricht dem wahren Parameterwert. 0.5 P

2. $\widehat{Var(H)} = \frac{1}{n-1}\sum_{i=1}^{n}(H_i - \overline{H}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}H_i^2 - \overline{H}_n^2\right) = \frac{100}{99}\left(\frac{190.43}{100} - \left(\frac{102.30}{100}\right)^2\right) =$ 0.8664 1.5 P

3. nicht eindeutig; asymptotisch erwartungstreu; MSE-konsistent; asymptotisch normalverteilt 0.5 P

4. a) $P(H = 3) = 0$ 0.5 P

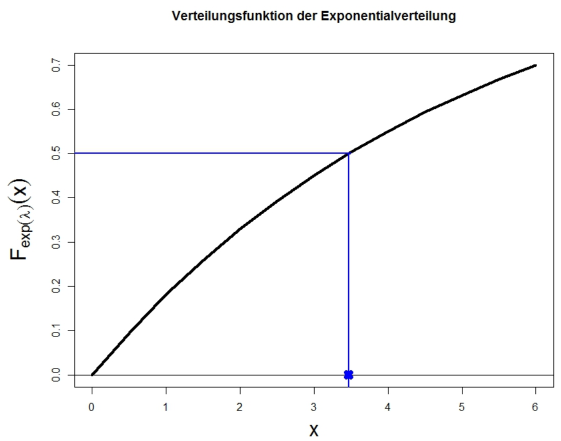
   b) $P(H > 1) = 1 - P(H \leq 1) = 1 - (1 - e^{-1}) = e^{-1} \approx 0.3679$ 1 P

   c) $P(H \leq 4|H > 3) = F_{Exp(1)}(1) = 1 - \exp\{-1\} = 0.6321$ 1 P

   oder: $P(H \leq 4|H > 3) = \frac{P(3 < H \leq 4)}{P(H > 3)} = \frac{F_{Exp(1)}(4) - F_{Exp(1)}(3)}{1 - F_{Exp(1)}(3)} = 0.6321$

5. $\lambda = \frac{1}{5} = 0.2$ 0.5 P

6. Der Median entspricht dem 50%-Quantil der Verteilung: 0.5 P




7. a) $\left[\frac{\chi^2_{\alpha;2n}}{2n\overline{Z}}\ ;\ \infty\right) = \left[\frac{\chi^2_{0.05;40}}{2\cdot 51.75}\ ;\ \infty\right) = \left[\frac{26.51}{103.5}\ ;\ \infty\right) = [0.2561\ ;\ \infty)$ 1.5 P

   b) In 95 von 100 Fällen liegt der wahre Parameter im real. Konfidenzintervall.

   **oder**

   Mit einer Vertrauenswürdigkeit von 95% liegt der Parameter im realisierten Konfidenzintervall 0.5 P

8. a) $t_{20} = 2 \cdot \frac{1}{5} \cdot 51.75 = 20.7$ 0.5 P

   b) $p = P(T_N \leq t_{20}) = P(T_N \leq 20.7) = 0.005 = 0.5\%$ 0.5 P

   $(p = P(T_N \leq t_{20}) = P(T_N \leq 22.16) = 1\%$

c) $p = 0.005(0.1) < 0.05 = \alpha$

$\Longrightarrow$ Nullhypothese kann auf dem 5%-Niveau abgelehnt werden,

**oder**

$\lambda_0 = 0.2$ liegt nicht im realisierten Konfidenzintervall, daher kann die Nullhypothese abgelehnt werden,

**oder**

$t_{20} = 20.7 < 26.51 = \chi^2_{0.05;40} = \chi^2_{\alpha;2n}$, daher kann die Null- hypothese abgelehnt werden. 1 P

# Lösung 3

1. a) $[\mu_w - 2 \cdot \sigma_k; \mu_w - 2 \cdot \sigma_k]$
   $= \quad [6.37 - 2 \cdot \sqrt{0.04}; 6.37 + 2 \cdot \sqrt{0.04}] = [5.97; 6.77]$ 1.0 P

   b) $\Phi(1.53) - \Phi(-1.53) = 2 \cdot 0.9370 - 1 = 0.8740$ 1.0 P

   c) $P(W \geq 6.50) = 1 - P(\frac{(W-\mu_W)}{\sigma_W} \leqslant \frac{(6.50-\mu_W)}{\sigma_W}) =$
   $1 - P(\frac{(W-\mu_W)}{\sigma_W} \leqslant \frac{(6.50-6.37)}{\sqrt{0.04}}) = 1 - \Phi(0.65) = 1 - 0.7422 = 0.2578$ 1.5 P

   d) $\bar{w}_n = 6.3480$ 0.5 P

   e) Spannweite $= 6.51 - 6.18 = 0.33$ 0.5 P

2. a) Likelihoodfunktion:
   $L = \Pi_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-(x_i-\mu)^2}{2\sigma^2}\}$
   Log-Likelihoodfunktion:
   $\ln(L) = n \; \ln(1) - n \; \ln(\sigma\sqrt{2\pi}) - (\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2})$
   $= -n \; \ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}$ 1.5 P

   b) 1. Ableiten: $\frac{\partial lnL}{\partial \mu} : \frac{\sum_{i=1}^n (x_i-\mu)}{\sigma^2}$ 0.5 P
   2. Gleich 0 setzen: $\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i$ 0.5 P

   c) $\frac{179}{20} = 8.95[\text{sec}]$ 0.5 P

   d) Die zweite Ableitung der (logarithmierten) Likelihoodfunktion muss an der Stelle des ML-Schätzers $\hat{\mu}_{ML}$ negativ sein. 0.5 P +0.5 P

3. a) $f_Z(z) = \frac{1}{43} = 0.0233$ 0.5 P

   b) Geometrische Verteilung 0.5 P

   c) $E[Z] = 42$ 0.5 P

# Lösung 4

1. a)

| | $F$ | $\overline{F}$ | $\Sigma$ |
|---|---|---|---|
| $W_1$ | 0.006 | 0.194 | 0.2 |
| $W_2$ | 0.0056 | 0.7944 | 0.8 |
| $\Sigma$ | 0.0116 | 0.9884 | 1 |

2 P

   b) 98.84% ✓ — 0.5 P

   c) 100% ✓ — 0.5 P

   d) $\frac{0.0056}{0.0116}$ ✓ $= 0.4828$ ✓ — 1 P

2. Sie stimmen überein.✓ — 0.5 P

3. $\sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n,p)$✓✓ — 1 P

4. $n \geq \left(\lambda_{1-\alpha/2}\frac{1}{2\varepsilon}\right)^2 = \left(1.6448\frac{1}{2\cdot 0.02}\right)^2$ ✓ $= 1690.854$ ✓d.h. $n \geq 1691$ ✓ — 1.5 P

5. a) 0.5 ✓ — 0.5 P

   b)

$$\begin{aligned} f(p) = p(1-p) \quad &\Rightarrow \quad f'(p) = 1 - 2p \;\checkmark \\ p \in\, ]0, 0.1] \quad &\Rightarrow \quad f'(p) \in\; [0.8, 1[ \text{ insbesondere: } f'(p) > 0 \;\checkmark \\ p \in\, ]0, 0.1] \quad &\Rightarrow \quad p(1-p) \leq 0.1(1-0.1) = 0.09 \;\checkmark \end{aligned}$$

1.5 P

   c) $n \geq \frac{\frac{9}{100}}{\alpha\varepsilon^2}$✓$= \frac{0.09}{0.1\cdot 0.02^2} = 2250$✓ — 1 P

# Klausur Statistik (10 ECTS)

| | | | |
|---|---|---|---|
| **Name** | | **Prüfer** | Prof. Dr. I. Klein |
| **Vorname** | | **Arbeitszeit** | Freitag, 19.02.2016 |
| **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 seit dem WS 2014/15 offiziell herausgegebene Formelsammlung (DIN A5, gebunden, orangener Umschlag). Es sind prinzipiell keine weiteren Eintragungen oder Markierungen darin erlaubt. Ausgenommen sind farbliche Hinterlegungen von Textpassagen und/oder Formeln. Außerdem ist das Erratum zur 1. Auflage der Formelsammlung 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

Anlässlich der Fußball-Europameisterschaft fahren die begeisterten Fußball-Fans Lukas und Basti mit dem Fanbus nach Frankreich. Sie vergleichen die beiden Länder Deutschland und Frankreich. In angegebener Tabelle sind ihre Ergebnisse festgehalten.[1]

| | Deutschland | Frankreich |
|---|---|---|
| Hauptstadt | Berlin | Paris |
| Bevölkerung in Tausend | 81174 | 66352 |
| Fläche in km$^2$ | 357137 | 632834 |
| Bruttoinlandsprodukt (BIP) je Einwohner in EUR | 36000 | 32200 |

1. Nennen Sie den höchsten Skalentyp des Merkmals 'Hauptstadt'.

2. Die Bevölkerung in Deutschland steigt im nächsten Jahr um 2% und sinkt im übernächsten Jahr um 2%. Wie groß ist die Bevölkerungszahl in zwei Jahren?

3. Wie hoch ist das Bruttoinlandsprodukt je Einwohner für beide Länder zusammen?

4. Der Bus fährt den ersten Streckenabschnitt, welcher eine Länge von 350 km besitzt, mit einer Durchschnittsgeschwindigkeit von 100 km/h. Der zweite Abschnitt beträgt 400 km und wird mit einer Durchschnittsgeschwindigkeit von 80 km/h zurückgelegt. Berechnen Sie die durchschnittliche Geschwindigkeit auf der gesamten Strecke.

[1] https://www.destatis.de/Europa/DE/Staat/EUStaaten/Deutschland.html , https://www.destatis.de/Europa/DE/Staat/EUStaaten/Frankreich.html

Schiedsrichter Walter Streng zieht im Mittel 4.5 Karten während eines Fußballspiels, wobei nicht zwischen gelben und roten Karten unterschieden werden soll. Die Anzahl der pro Spiel gezogenen Karten soll durch die Poisson-verteilte Zufallsvariable $Y$ beschrieben werden. Im Folgenden werden nur Fußballspiele mit einer Dauer von 90 Minuten betrachtet, das heißt es gibt keine Verlängerung, kein Elfmeterschießen oder Ähnliches.

5. Welchen Wert nimmt der Parameter für die Verteilung von $Y$ an?

**Arbeiten Sie jetzt mit $\underline{\boldsymbol{\lambda = 3}}$ weiter!**

6. Mit welcher Wahrscheinlichkeit zeigt der Schiedsrichter in einem Spiel genau 2 Karten?

7. Berechnen Sie die Wahrscheinlichkeit, dass der Schiedsrichter im ersten Spiel mehr als 2 Karten und im zweiten Spiel höchstens 3 Karten vergibt. Gehen Sie hierbei davon aus, dass die Anzahl der Karten im ersten Spiel ($Y_1$) unabhängig von der Anzahl der Karten im zweiten Spiel ($Y_2$) ist.

8. Die Zufallsvariable $W$ beschreibt die Wartezeit zwischen zwei Ereignissen in Minuten. Nationalspieler Thomas Nüller hat gerade die gelbe Karte gesehen. Berechnen Sie die Wahrscheinlichkeit, dass innerhalb der nächsten 10 Minuten ein beliebiger Spieler die nächste Karte sieht.

Der Trainer hat die Hypothese, dass all seine 23 Spieler eine Schussgeschwindigkeit haben, deren Varianz nicht größer als 64 ist. Sie testen die Hypothese mit einem geeigneten Hypothesentest, wobei die Irrtumswahrscheinlichkeit 5% beträgt. Dafür betrachten Sie die Schussgeschwindigkeiten der 23 Spieler $X_1, X_2, \ldots, X_{23}$ als stochastisch unabhängig, identisch normalverteilt. Sie stellen folgende Hypothesen auf:

$$H_0 : \sigma^2 \leq \sigma_0^2 = 64 \qquad \text{gegen} \qquad H_1 : \sigma^2 > \sigma_0^2 = 64.$$

9. Zeigen Sie, dass die folgende Gleichung gilt: $\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2$.

10. Bestimmen Sie die realisierte Prüfgröße, wenn für die Stichprobe $\sum_{i=1}^{23} x_i^2 = 309570$ und $\overline{x}_{23} = 116$ gelten.

**Verwenden Sie im Folgenden als Wert für die realisierte Prüfgröße 5 und als Wert für die kritische Schranke 40!**

11. Treffen Sie die Testentscheidung und begründen Sie diese.

**Schmierpapier zu Aufgabe 1**

# Aufgabe 2

Nach der Überschwemmung der Stadt P vor zwei Jahren versuchen Sie den Stadtrat vom Erwerb eines mobilen Hochwasserschutzes zu überzeugen. Sei

$H$ : „Höhe des Hochwassers, falls der Fluss D sein Flussbett verlässt" [m].

Aus einer einfachen Stichprobe $H_1, ..., H_{100}$ aus den letzten 100 Jahre haben Sie sich bereits folgende Daten extrahiert:

$$\sum_{i=1}^{100} h_i = 102.30 \qquad \text{und} \qquad \sum_{i=1}^{100} h_i^2 = 190.43.$$

1. Was bedeutet „erwartungstreu"?

2. Geben Sie einen erwartungstreuen Schätzer für die Varianz an und berechnen Sie den zugehörigen Schätzwert.

3. Nennen Sie eine Eigenschaft eines Momenten-Schätzers.

Sie wissen, dass

$$P(H \leq x) = 1 - \exp(-x) \quad \text{mit zugehöriger Dichtefunktion} \quad f(x) = \exp(-x)$$

für $x \geq 0$.

4. Der Stadtrat behauptet, dass die Wahrscheinlichkeit, dass das Hochwasser die gängigen Barrikaden überschreitet, sehr gering sei. Berechnen Sie

   a) die Wahrscheinlichkeit, dass der Hochwasserstand genau 3 m beträgt,

   b) die Wahrscheinlichkeit, dass der Hochwasserstand eine Barrikade von 1 m überschreitet,

   c) die Wahrscheinlichkeit, dass der Hochwasserstand maximal 4 m beträgt, wenn er die 3 m Marke bereits überschritten hat.

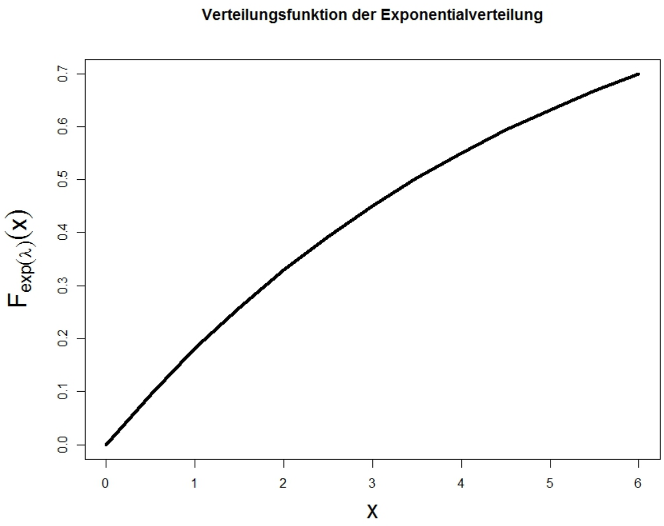
Sie drängen den Stadtrat zu einer schnellen Entscheidung, da ihrer Meinung nach der Fluss D im Mittel alle 5 Jahre eine Überschwemmung verursacht. Sie wissen, dass

$Z$ : „Zeit zwischen zwei Überschwemmungen in Jahren"

exponentialverteilt ist mit Parameter $\lambda$.

5. Geben Sie auf Grundlage der erhaltenen Informationen den Parameter $\lambda$ an.

6. Markieren Sie den Median der Exponentialverteilung in der nachstehenden Graphik eindeutig.




Ein Mitglied des Stadtrates widerspricht Ihnen und äußert die Vermutung, dass der Fluss D seltener über die Ufer tritt. Sie betrachten hierzu eine einfache Stichprobe $Z_1, ..., Z_{20}$ mit $\sum_{i=1}^{20} z_i = 51.75$.

7. a) Geben Sie das theoretische und das realisierte **rechtsseitige** Konfidenzintervall für den Parameter $\lambda$ zum Niveau $1 - \alpha = 0.95$ an.

b) Interpretieren Sie das realisierte Konfidenzintervall.

8. Testen Sie nun auf dem 95%-Niveau Ihre Vermutung gegen die Ihres Kontrahenten in Form der Hypothesen

$$H_0: \ \lambda \leq \frac{1}{5} = \lambda_0 \qquad \text{gegen} \qquad H_1: \ \lambda > \frac{1}{5}$$

mit zugehöriger Prüfgröße

$$T_n = 2\lambda_0 \sum_{i=1}^{n} Z_i \sim \chi^2_{2n}$$

a) Berechnen Sie die realisierte Prüfgröße $t_{20}$.

b) Berechnen Sie den p-Wert. **Sollten Sie in Aufgabe 8a) keinen Wert erhalten haben, verwenden Sie $t_{20} = 22.16$.**

c) Treffen Sie eine Testentscheidung und begründen Sie diese in Form einer Entscheidungsregel.

**Schmierpapier zu Aufgabe 2**

# Aufgabe 3

Ihr Freund Kevin ist begeisterter Sportler. Sie besuchen Kevin bei seinem Training.

1. Zuerst beobachten Sie den Weitsprung. Nehmen Sie zunächst an, dass beim Weitsprung-Training im Allgemeinen die Ergebnisse der gesprungenen Weiten (Zufallsvariable W) in Meter unabhängig und identisch normalverteilt sind mit $W \sim N(\mu = 6.37, \sigma^2 = 0.04)$.

   a) Berechnen Sie die Grenzen des zweifachen zentralen Schwankungsintervalls für die Sprungweite W.

   b) Mit welcher Wahrscheinlichkeit liegt die gesprungene Weite eines Sportlers im 1.53-fachen zentralen Schwankungsintervall?

   c) Mit welcher Wahrscheinlichkeit springt ihr Freund Kevin mindestens eine Weite von 6.50 Metern?

   Nun notieren Sie sich die gesprungenen Weiten (W) in Meter der ersten 10 Sprünge von Kevin:

| **Sprung i** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $w_i$ | 6.30 | 6.29 | 6.51 | 6.22 | 6.49 | 6.27 | 6.46 | 6.18 | 6.40 | 6.36 |

   d) Berechnen Sie die durchschnittliche gesprungene Weite der 10 beobachteten Sprünge.

e) Berechnen Sie die Spannweite der beobachteten gesprungenen Weiten.

2. Als nächstes wenden Sie sich dem Training im Sprint zu und beobachten 20 Sprinter. Nehmen Sie an, die „Zeiten im 100m-Sprint [in Sekunden]" (X) sind normalverteilt mit bekannter Varianz $\sigma^2$.

   Es wird der Maximum-Likelihood-Schätzer für den unbekannten Parameter $\mu$ gesucht.

   a) Zeigen Sie, dass die logarithmierte Likelihoodfunktion von $\mu$ gegeben ist durch

   $$\ln(L(\mu; X_1, ..., X_n)) = -n \ \ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2}.$$

   b) Zeigen Sie, dass der Maximum-Likelihood-Schätzer für den unbekannten Parameter $\mu$ lautet: $\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

   c) Berechnen Sie den Wert von $\hat{\mu}_{ML}$, wenn Sie wissen, dass $\sum_{i=1}^{20} x_i = 179$.

d) Formulieren Sie in wenigen Worten die hinreichende Bedingung für das Vorliegen eines globalen Maximums an der Stelle $\hat{\mu}_{ML}$.

Zum Abschluss seines heutigen Trainings macht Kevin noch ein paar Klimmzüge. Sie beobachten ihn und stellen fest, dass die Klimmzugstange schon sehr alt ist.

3. Nehmen Sie an, die Wahrscheinlichkeit, dass die Klimmzugstange zerbricht, sei bei jedem einzelnen Klimmzug gleichbleibend $p = \frac{1}{43}$. Außerdem besitze die Anzahl der Klimmzüge (Z) bevor die Klimmzugstange zerbricht folgende Wahrscheinlichkeitsfunktion:

$$f_Z(z) = p(1-p)^z.$$

a) Wie hoch ist die Wahrscheinlichkeit, dass die Stange bei Kevins erstem Klimmzug zerbricht?

b) Wie nennt sich die Verteilung von Z?

c) Wie viele Klimmzüge sind zu erwarten, bevor die Klimmzugstange zerbricht?

**Schmierpapier zu Aufgabe 3**

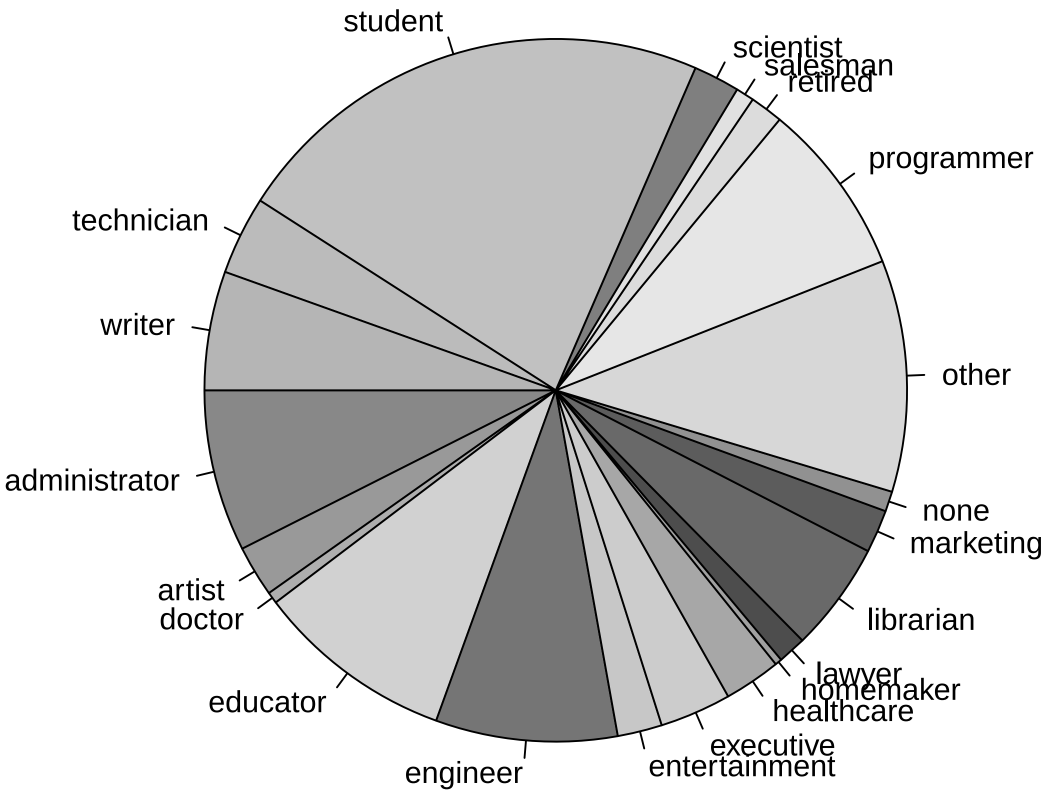
# Aufgabe 4

In Ihrem Workspace liegt ein bereits erstellter Dataframe `MovieLense` mit Daten aus dem *MovieLense 100 K Datensatz*[2] vor, in welchem die Nutzerbewertungen von Filmen aufgelistet sind. Weiterhin sind in ihm folgende Merkmale enthalten:

| **Merkmal** | | **Spaltenname** |
|---|---|---|
| *User ID*: | Identifikationsnummer des Users | `user_id` |
| *Filmtitel*: | Titel des bewerteten Films | `movie_title` |
| *Genre*: | Genre des bewerteten Films | `genre` |
| *Rating*: | Bewertung des Films (1 bis 5 Sterne) | `rating` |
| *Alter*: | Alter des Users | `age` |
| *Geschlecht*: | Geschlecht des Users | `gender` |
| *Beruf*: | Beruf(sgruppe) des Users | `occupation` |

Die folgende Grafik enthält ein Tortendiagramm, welches erstellt wurde durch den Befehl:

```
> pie(table(MovieLense$occupation))
```




[2]https://github.com/JustinChu/STAT545A_MovieStats.git

1. Welche Merkmalsausprägung bildet den Modus des Merkmals *Beruf*? Ist die Modalhäufigkeit größer, kleiner oder gleich 33%?

Folgender R-Code (*Code 1*) wird nun in die Konsole eingegeben:

```
> str(MovieLense[,c("genre","gender","age","rating")])
'data.frame':        212257 obs. of  4 variables:
 $ genre : Factor w/ 18 levels "Action","Adventure",..: 4 5 5 4 4 5 5 4 5 4 ...
 $ gender: Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 1 1 ...
 $ age   : int  24 24 15 15 47 47 49 49 49 49 ...
 $ rating: int  2 2 3 3 2 2 3 3 3 3 ...
> class(MovieLense$occupation)
[1] "factor"
> f = table(MovieLense$occupation)/length(MovieLense$occupation)
> ge = sum(f*(1-f))
> ge2 = 1-sum(f^2)
```

2. Geben Sie alle statistischen Kenngrößen an, welche durch *Code 1* im Workspace abgespeichert werden.

3. Lässt sich etwas über das Größenverhältnis zwischen `ge` und `ge2` aussagen? (**Kurze** Begründung)

4. Geben Sie R-Befehle an, um die Lagemaße der Merkmale *Alter*, *Geschlecht* und *Rating* zu erhalten, welche für die jeweiligen Merkmalstypen am besten geeignet sind.

Sie werden nun beauftragt das Rating-Verhalten für gewisse Zielgruppen zu analysieren.

5. Geben Sie den `R`-Code an, um sich ausschießlich Berufe der User anzeigen zu lassen, welche Filme des Genres *Action* bewertet haben.

6. Mit welchem `R`-Code können Sie sich den Teildatensatz anzeigen lassen, der lediglich die Zielgruppe der 18 bis 30-jährigen User enthält?

100 neue User aus der gleichen Grundgesamtheit möchten ebenfalls Filme bewerten. Bevor sie das können, sollen sie an einer Umfrage teilnehmen, bei der u.A. das Merkmal *Geschlecht* abgefragt wird. Die Antworten wurden mit den **Zahlen** '0' für die Ausprägung *männlich* und '1' für die Ausprägung *weiblich* codiert und liegen bereits als Vektor `geschlecht` im Workspace vor. Folgender Befehl wird in die Konsole eingegeben:

```
> class(geschlecht)

[1] "character"
```

7. Geben Sie den `R`-Code an, mit dem Sie die notwendige Transformation des Vektors `geschlecht` durchführen können.

Die nötige Transformation sei nun durchgeführt. Es folgen die Befehle (*Code 2*):

```
> p0 = 0.25
> c(qnorm(0.17),qnorm(0.085,mean(geschlecht),sd(geschlecht)),qnorm(0.085))

[1] -0.9541653 -0.3392219 -1.3722038

> sqrt(length(geschlecht))*(mean(geschlecht)-p0)/sqrt(p0*(1-p0))

[1] 0.6928203
```

8. Verwenden Sie *Code 2* um (approximativ) die Hypothese für den Anteilswert $H_0 : p \geq p_0 = 0.25$ bei einer Irrtumswahrscheinlichkeit von $\alpha = 17\%$ unter Angabe der notwendigen Zwischenschritte zu testen.

9. Formulieren Sie in wenigen Worten die Vermutung, welche zu dem Hypothesenpaar in Teilaufgabe 8. führt.

10. Der $p$-Wert des Tests aus Teilaufgabe 8. für eine Stichprobe $(x_1, ..., x_n)$ wird mit der Funktion

$$\Phi\left(\sqrt{n}\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\right)$$

berechnet. Vervollständigen Sie den nachstehenden `R`-Code, damit die Funktion `p.value` zur Berechnung des $p$-Wertes für die Stichprobe `geschlecht` verwendet werden kann.

```
p.value = function(geschlecht,p0,n){


}
```

**Schmierpapier zu Aufgabe 4**

# Lösung Klausur WS1516 (10 ECTS)

## Lösung 1

1) Nominalskala 0.5 P

2) $81174 \cdot 1.02 \cdot 0.98 = 81141.5304$ 0.5 P

3) $\dfrac{36000 \cdot 81174 + 32200 \cdot 66352}{81174 + 66352} = 34290.8938$ 1.0 P

4) $\dfrac{350 + 400}{\frac{350}{100} + \frac{400}{80}} = 88.2353$ 1.0 P

5) $\lambda = 4.5$ 0.5 P

6) $P(Y_1 = 2) = 0.2240$ 0.5 P

7) $P(Y_1 > 2)P(Y_2 \leq 3) = (1 - 0.4232) \cdot 0.6472 = 0.3733$ 1.5 P

8) $F_{exp}(10; \frac{3}{90}) = 1 - exp(-\frac{3}{90} \cdot 10) = 0.2835$ 1.5 P

9) $\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum_{i=1}^{n} x_i^2 - \underbrace{\sum_{i=1}^{n} 2x_i\,\overline{x}_n}_{=2n\overline{x}_n} + \underbrace{\sum_{i=1}^{n} \overline{x}_n^2}_{=n\overline{x}_n^2} = \sum_{i=1}^{n} x_i^2 - 2n\overline{x}_n^2 + n\overline{x}_n^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2$ 1.0 P

10) $t_n = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}{\sigma_0^2} = \dfrac{309570 - 23 \cdot 116^2}{64} = 1.2813$ 1.0 P

11) Lehne $H_0$ nicht ab, da $PG < KS$ (PG nicht im KB). 1.0 P

# Lösung 2

1. Der Erwartungswert des Schätzers entspricht dem wahren Parameterwert. 0.5 P

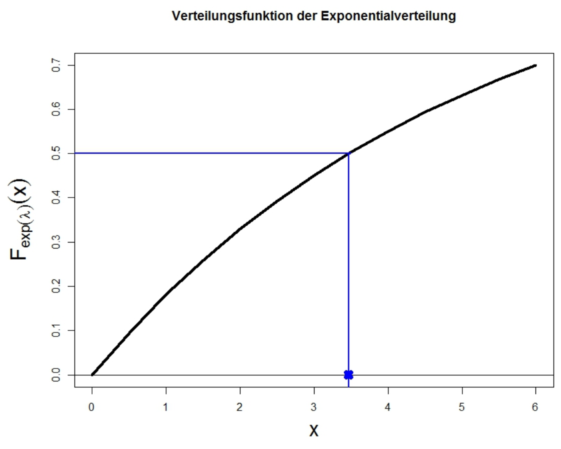
2. $\widehat{Var(H)} = \frac{1}{n-1}\sum_{i=1}^{n}(H_i - \overline{H}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}H_i^2 - \overline{H}_n^2\right) = \frac{100}{99}\left(\frac{190.43}{100} - \left(\frac{102.30}{100}\right)^2\right) =$ 0.8664 1.5 P

3. nicht eindeutig; asymptotisch erwartungstreu; MSE-konsistent; asymptotisch normalverteilt 0.5 P

4. a) $P(H = 3) = 0$ 0.5 P
   b) $P(H > 1) = 1 - P(H \leq 1) = 1 - (1 - e^{-1}) = e^{-1} \approx 0.3679$ 1 P
   c) $P(H \leq 4 | H > 3) = F_{Exp(1)}(1) = 1 - \exp\{-1\} = 0.6321$ 1 P
   oder: $P(H \leq 4|H > 3) = \frac{P(3 < H \leq 4)}{P(H > 3)} = \frac{F_{Exp(1)}(4) - F_{Exp(1)}(3)}{1 - F_{Exp(1)}(3)} = 0.6321$

5. $\lambda = \frac{1}{5} = 0.2$ 0.5 P

6. Der Median entspricht dem 50%-Quantil der Verteilung: 0.5 P




7. a) $\left[\frac{\chi^2_{\alpha;2n}}{2n\overline{Z}}\ ;\ \infty\right) = \left[\frac{\chi^2_{0.05;40}}{2 \cdot 51.75}\ ;\ \infty\right) = \left[\frac{26.51}{103.5}\ ;\ \infty\right) = [0.2561\ ;\ \infty)$ 1.5 P
   b) In 95 von 100 Fällen liegt der wahre Parameter im real. Konfidenzintervall.
   **oder**
   Mit einer Vertrauenswürdigkeit von 95% liegt der Parameter im realisierten Konfidenzintervall 0.5 P

8. a) $t_{20} = 2 \cdot \frac{1}{5} \cdot 51.75 = 20.7$ 0.5 P
   b) $p = P(T_N \leq t_{20}) = P(T_N \leq 20.7) = 0.005 = 0.5\%$ 0.5 P
   $(p = P(T_N \leq t_{20}) = P(T_N \leq 22.16) = 1\%$

c) $p = 0.005(0.1) < 0.05 = \alpha$

$\Longrightarrow$ Nullhypothese kann auf dem 5%-Niveau abgelehnt werden,

**oder**

$\lambda_0 = 0.2$ liegt nicht im realisierten Konfidenzintervall, daher kann die Nullhypothese abgelehnt werden,

**oder**

$t_{20} = 20.7 < 26.51 = \chi^2_{0.05;40} = \chi^2_{\alpha;2n}$, daher kann die Null- hypothese abgelehnt werden. 1 P

# Lösung 3

1. a) $[\mu_w - 2 \cdot \sigma_k; \mu_w - 2 \cdot \sigma_k]$
   $= \quad [6.37 - 2 \cdot \sqrt{0.04}; 6.37 + 2 \cdot \sqrt{0.04}] = [5.97; 6.77]$ 1.0 P

   b) $\Phi(1.53) - \Phi(-1.53) = 2 \cdot 0.9370 - 1 = 0.8740$ 1.0 P

   c) $P(W \geq 6.50) = 1 - P(\frac{(W-\mu_W)}{\sigma_W} \leqslant \frac{(6.50-\mu_W)}{\sigma_W}) =$
   $1 - P(\frac{(W-\mu_W)}{\sigma_W} \leqslant \frac{(6.50-6.37)}{\sqrt{0.04}}) = 1 - \Phi(0.65) = 1 - 0.7422 = 0.2578$ 1.5 P

   d) $\bar{w}_n = 6.3480$ 0.5 P

   e) Spannweite $= 6.51 - 6.18 = 0.33$ 0.5 P

2. a) Likelihoodfunktion:
   $L = \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-(x_i-\mu)^2}{2\sigma^2}\}$
   Log-Likelihoodfunktion:
   $\ln(L) = n\ \ln(1) - n\ \ln(\sigma\sqrt{2\pi}) - (\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2})$
   $= -n\ \ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$ 1.5 P

   b) 1. Ableiten: $\frac{\partial lnL}{\partial\mu} : \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}$ 0.5 P
   2. Gleich 0 setzen: $\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 0.5 P

   c) $\frac{179}{20} = 8.95[\text{sec}]$ 0.5 P

   d) Die zweite Ableitung der (logarithmierten) Likelihoodfunktion muss an der Stelle des ML-Schätzers $\hat{\mu}_{ML}$ negativ sein. 0.5 P +0.5 P

3. a) $f_Z(z) = \frac{1}{43} = 0.0233$ 0.5 P

   b) Geometrische Verteilung 0.5 P

   c) $E[Z] = 42$ 0.5 P

# Lösung 4

1. Student mit Häufigkeit kleiner als 33% — 0.5 P + 0.5 P

2. Relative Häufigkeiten des Mermals `occupation`, Gini Entropie (`ge` und `ge2`)
   0.5 P + 0.5 P

3. Ja – identisch, da lediglich alternative Berechnung — 0.5 P + 0.5 P

4. `mean(MovieLense$age)`, `which.max(table(MovieLense$gender))` und `quantile(MovieLense$rating,0.5,type = 1)` — 0.5 P + 0.5 P + 0.5 P

5. `MovieLense[MovieLense$genre == "Action","occupation"]`
   1 P (-0.5 P pro Fehler)

6. `MovieLense[MovieLense$age > 17 & MovieLense$age < 31,]`
   1 P (-0.5 P pro Fehler)

7. `geschlecht = as.numeric(geschlecht)` — 0.5 P

8. $t_n = 0.6928203 \not< -0.9541653 = -\lambda_{0.83}$, $H_0$ ablehnen bei $\alpha = 0.17\%$
   0.5 P + 0.5 P + 0.5 P

9. Anteil an weiblichen Benutzern ist geringer als 25%. — 0.5 P

10. 
```
p.value = function(geschlecht,p0,n){
          pnorm(sqrt(n)*(mean(geschlecht)-p0)/sqrt(p0*(1-p0)))
}
```
   0.5 P (`pnorm()`) + 0.5 P korrekt