Klausur Statistik (7.5 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Freitag, 11.08.2017
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

Sie waren am Wochenende mit Ihren Freunden Kartfahren und haben sieben Runden absolviert. Sie betrachten die Zufallsvariable X: „Zeit pro Runde (in Sekunden)". Gehen Sie davon aus, dass gilt $X \sim N(\mu_X, \sigma_X^2)$.

1. Wie groß ist die Wahrscheinlichkeit in der 8. Runde eine Zeit von genau $42,\overline{3}$ Sekunden zu erreichen?

2. Sie berechnen das Stichprobenmittel als $\bar{x}_7 = 42.2286$. Wie würde sich das Stichprobenmittel tendenziell verändern, wenn Sie aufgrund eines technischen Defekts für die dritte Runde 70 Sekunden statt 41.7 Sekunden gebraucht hätten? (Hinweis: Eine Berechnung ist nicht notwendig.)

3.(a) Sie haben für einen allgemeinen Stichprobenumfang n und beliebiges $\epsilon > 0$ folgende Formel gegeben:

$$P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) = 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma_X}\right) - 1$$

Zeigen Sie mit Hilfe dieser Formel: $\lim_{n\to\infty} P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) = 1$.

(b) Ist das Stichprobenmittel $\bar{X}_n$ ein konsistenter Schätzer für μ_X? Begründen Sie Ihre Entscheidung.

4. Berechnen Sie mit Hilfe von $\sum_{i=1}^{7} x_i^2 = 12490.66$ und $\bar{x}_7 = 42.2286$ die Stichprobenvarianz.

5. Gehen Sie nun von einer Stichprobenvarianz von $s_7^2 = 1.3$ Sek^2 aus. Sie wollen bei einer Irrtumswahrscheinlichkeit von 5% folgendes Hypothesenpaar testen:

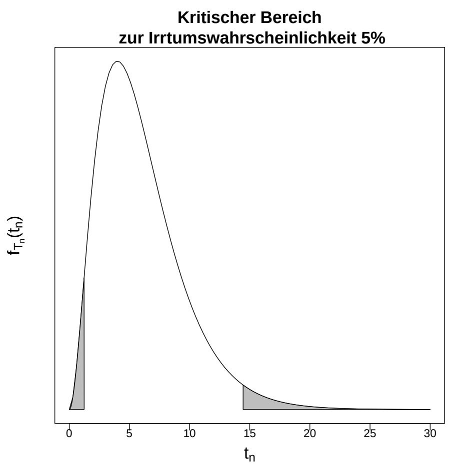
$$H_0 : \sigma^2 = 1 \text{ vs. } H_1 : \sigma^2 \neq 1$$

(a) Geben Sie die Prüfgröße sowie deren theoretische Verteilung an und berechnen Sie den Wert der realisierten Prüfgröße.

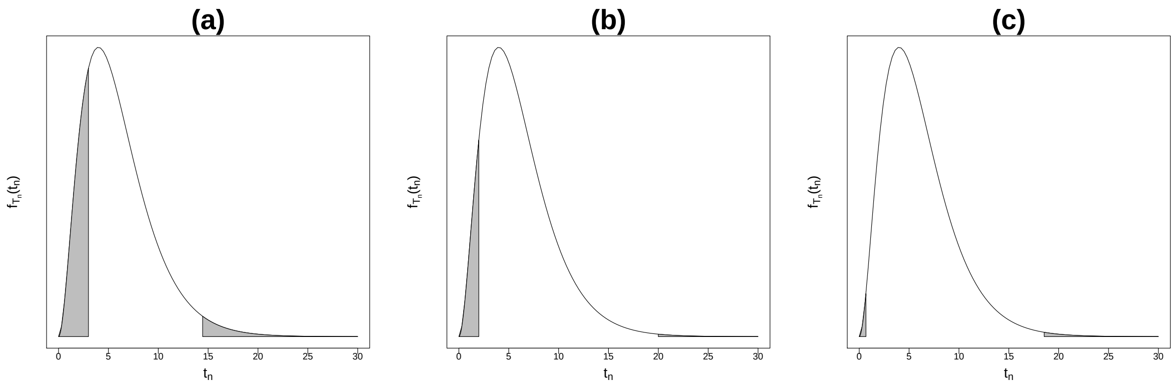
(b) Geben Sie die kritischen Schranken des Tests an.

(c) Treffen Sie eine Testentscheidung und formulieren Sie einen Antwortsatz. (Hinweis: Falls Sie in Teilaufgabe (b) bzw. (c) kein Ergebnis erhalten haben, benutzen Sie für die untere Schranke des kritischen Bereichs den Wert 3, für die obere Schranke des kritischen Bereichs den Wert 18 und für die realisierte Prüfgröße den Wert 4.)

(d) Im Folgenden bezeichne f die Dichtefunktion unter der Nullhypothese und $T_n = (n-1)S_n^2$. Grafisch lässt sich der kritische Bereich zu einer Irrtumswahrscheinlichkeit von 5% wie folgt darstellen:



Angenommen Sie hätten eine Irrtumswahrscheinlichkeit von 1% festgelegt. Welche der folgenden drei Grafiken würde sich dann ergeben?



Nehmen Sie nun an, die wahre Verteilung der Zufallsvariable ist gegeben durch:

$$X \sim N(42, 1)$$

6. Berechnen Sie die Wahrscheinlichkeit in der 8. Runde eine Zeit zwischen 41.8 und 42.5 Sekunden zu erreichen.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

Sie suchen eine Wohnung in Nürnberg in Nähe des Stadtzentrums.

1. Nach längerer Suche erhalten Sie folgende Aufstellung der Anzahl von Wohnungen gegliedert nach Nähe zum Stadtzentrum und in welchem Medium Sie das Inserat fanden:

Medium	Nähe Stadtzentrum	Nicht in Nähe Stadtzentrum
Internetportal	5	2
Schwarzes Brett der Universität	1	1
Zeitungsanzeige	2	4

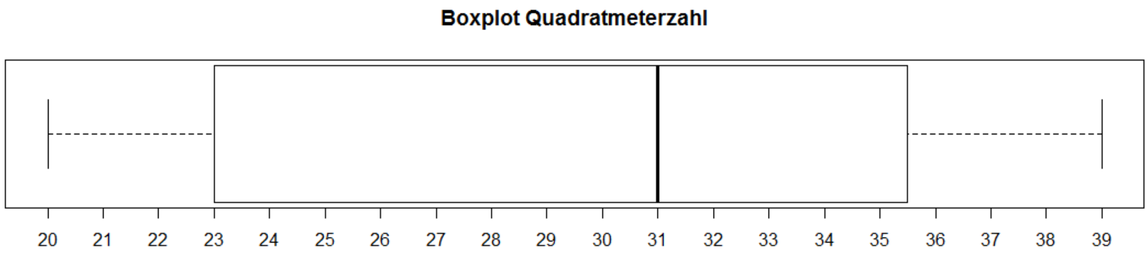
a) Welchen Merkmalstyp hat das Merkmal 'Medium'?

b) Wie viele Inserate haben Sie in der Tabelle gelistet?

c) Berechnen Sie den Anteil der Wohnungen, die nicht in Nähe des Stadtzentrums liegen.

d) Berechnen Sie den Anteil der Wohnungen, die in Stadtnähe liegen und die Sie als Zeitungsanzeige gefunden haben.

2. Sie listen nun die Wohnfläche [in qm] (X) der betrachteten Wohnungen auf und visualisieren die Daten mit folgendem Boxplot:



Der Modus liegt bei $x = 35$.

a) Geben Sie das untere Quartil ($x_{0.25}$) der Verteilung der Wohnfläche an.

b) Geben Sie den Median (x_{med}) der Verteilung der Wohnfläche an.

c) Ist die unimodale Verteilung der Wohnfläche rechtsschief, linksschief oder symmetrisch? Begründen Sie Ihre Antwort.

Nehmen Sie nun und im Folgenden an, die Verteilung der Wohnfläche folge einer Normalverteilung mit Mittelwert $\mu = 27$ und Varianz $\sigma^2 = 59$.

d) Berechnen Sie den ersten Pearsonschen Schiefekoeffizienten.

e) Geben Sie den Variationskoeffizienten an.

f) Wie viele Inserate müssen Sie betrachten, um mit einer Genauigkeit von $\pm$ 5 Quadratmetern und einer Irrtumswahrscheinlichkeit von 1% sagen zu können, wie groß die Wohnung, die sich sich leisten können, im Mittel sein wird?

Sei B: ′Besichtigte Wohnung besitzt einen Balkon′.

3. Von 27 Wohnungen die Sie besichtigen, haben 20% einen Balkon. Testen Sie Ihre Vermutung, dass die Wahrscheinlichkeit, dass eine Wohnung in Ihrer Preisklasse einen Balkon besitzt unter 30% liegt. Führen Sie einen Hypothesentest mit einer Irrtumswahrscheinlichkeit von 10% durch, wobei $B \sim Bernoulli(.)$.

a) Beschreiben Sie verbal den Fehler 1. Art.

b) Berechnen Sie die Prüfgröße des approximativen Tests.

c) Geben Sie die kritische Schranke für den Test an.

Hinweis: p-Wert= 0.1284.

d) Treffen Sie eine Testentscheidung und begründen Sie diese.

Nehmen Sie an, die Wahrscheinlichkeit, dass eine Wohnung einen Balkon hat, liegt bei 30%.

e) Berechnen Sie die Kurtosis der Verteilung von B.

Sollten Sie keine Lösung für Aufgabe 3.e) bekommen haben, verwenden Sie: $\kappa = -1.5$.

f) Weist der Wert der Kurtosis auf eine mesokurtische, leptokurtische oder platykurtische Verteilung hin?

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

Sie sind Qualitätsprüfer bei der Firma „Gear“, welche Getriebe produziert. Ihre Aufgabe ist es zu prüfen, ob die produzierten Getriebe defekt sind. Nehmen Sie an, die Zufallsvariable

X : „Anzahl der defekten Getriebe”

ist binomialverteilt mit den Parametern $n = 10$ und $p = 0.2$.

1. Nennen Sie das höchste Skalenniveau von X.

2. Berechnen Sie die Wahrscheinlichkeit, dass

   (a) genau zwei Getriebe defekt sind.

   (b) höchstens drei Getriebe defekt sind.

   (c) mindestens vier Getriebe nicht defekt sind.

   (d) mehr als ein Getriebe, aber weniger als fünf defekt sind.

3. Bestimmen Sie,

   (a) den Erwartungswert von X.

   (b) die Varianz von X.

   (c) den Modus von X.

4. Gehen Sie unabhängig von den vorherigen Ergebnissen davon aus, dass gilt:

   $E[X] = 1.2$ und $VAR[X] = 2$.

   Gegeben sei die Zufallsvariable $Y = a + bX$, mit $a, b \in \mathbb{R}$. Berechnen Sie

   (a) den Erwartungswert von Y.

   (b) die Varianz von Y.

Betrachten Sie die folgenden vier voneinander unabhängigen Zufallsvariablen:

$$
\begin{aligned}
W &\sim N(\mu_W, \sigma_W^2), \\
Z &\sim N(\mu_Z, \sigma_Z^2), \\
A &\sim N(0, 1), \\
B &\sim N(0, 1).
\end{aligned}
$$

5. Geben Sie die Namen der Verteilung, **inklusive** aller Parameter, der nachfolgenden Zufallsvariablen an:

   (a) $K = 3W$

   (b) $J = Z + A$

   (c) $H = A^2 + B^2$

6. Gegeben sei folgende nicht-negative Funktion, mit $\alpha > 0$:

$$f_X(x) = \begin{cases} \alpha(1+x^2) & \text{für} \quad -1 \leq x \leq 0, \\ \alpha x & \text{für} \quad 0 \leq x \leq 2, \\ 0 & \text{sonst.} \end{cases}$$

(a) Nennen Sie die zwei Eigenschaften die eine Dichtefunktion besitzen muss.

(b) Welchen Wert muss α annehmen, damit f_X als Dichtefunktion für eine Zufallsvariable X interpretiert werden kann?

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

Eine Fluggesellschaft überbucht ihr Flugzeug, da erfahrungsgemäß nicht alle Passagiere den Flug antreten. Es werden für einen Flug von Nürnberg nach Malta 92 Tickets verkauft, obwohl das Flugzeug nur 90 Plätze hat. Die Check-In Zeit ist zwischen 17:00 und 18:00 Uhr. Gäste, die nicht innerhalb der vorgesehenen Check-In Zeit erscheinen, können den Flug nicht antreten.
Sei

Y : „Anzahl der den Flug nicht antretenden Passagiere"

mit $Y \sim \text{Pois}(\lambda), \lambda = 2$.

1. Berechnen Sie die Wahrscheinlichkeit, dass

   (a) genau ein Gast nicht zum Flug erscheint.

   (b) höchstens 6 Gäste zum Flug nicht erscheinen.

   (c) alle den Flug antretenden Gäste einen Platz bekommen.

2. Die Wahrscheinlichkeit, dass Y eine Schwelle $y \in \{0, 1, 2, ...\}$ überschreitet beträgt 0.0011. Bestimmen Sie y.

3. Unter welchen Umständen kann die Binomialverteilung durch die Poissonverteilung approximiert werden und mit welchen Parametern?

4. Berechnen Sie die Unter- und Obergrenze des 4-fachen zentralen Schwankungsintervalls für Y.

5. Schätzen Sie mit der Ungleichung von Chebyshev die Wahrscheinlichkeit, dass Y im 4-fachen zentralen Schwankungsintervall liegt, nach unten ab.

6. Berechnen Sie die exakte Wahrscheinlichkeit, dass Y zwischen $\lambda - \sqrt{\lambda}$ und $\lambda + \sqrt{\lambda}$ liegt. Sie können hierbei folgende Zusammenhänge benutzen:
   $F_{Pois}(3.4142, \lambda = 2) = 0.8571, F_{Pois}(4, \lambda = 2) = 0.9473,$
   $F_{Pois}(3.4142, \lambda = 3) = 0.6472$ $F_{Pois}(0.58579, \lambda = 2) = 0.1353,$
   $F_{Pois}(1, \lambda = 2) = 0.4060, F_{Pois}(0.58579, \lambda = 3) = 0.0498.$

Die Wartezeit Z zwischen zwei zufällig am Check-In Schalter eintreffenden Passagieren sei exponentialverteilt mit Parameter $\lambda > 0$.

7. Zeigen Sie, dass die logarithmierte Likelihood Funktion der Exponential-Verteilung folgende Form hat:

$$LL(\lambda; X_1, ..., X_n) = n\log(\lambda) - \lambda \sum_{i=1}^{n} X_i$$

8. Zeigen Sie, dass der Maximum-Likelihood-Schätzer für λ folgende Form hat:

$$\hat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Gehen Sie davon aus, dass die hinreichenden Bedingungen für ein Maximum erfüllt sind.

Nun sei folgende Stichprobe aus 5 Wartezeiten in Minuten gegeben.

i	1	2	3	4	5
Wartezeit i	6.1	2.5	0.4	1.1	5.0

9. Rechnen Sie den ML-Schätzer $\hat{\lambda}_{ML}$ für die gegebene Stichprobe aus.

10. Welche Eigenschaften aus den folgenden erfüllt ein ML-Schätzer im Allgemeinen? Eindeutig, asymptotisch erwartungstreu, erwartungstreu.

Schmierpapier zu Aufgabe 4

Lösung Klausur SoSe 2017 (7.5 ECTS)

Lösung 1

1. $P(X = 42,\overline{3}) = 0$ 0.5 P

2. Steigt 0.5 P

3. (a)

$$\begin{aligned}\lim_{n\to\infty} P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) &= \lim_{n\to\infty} 2\Phi(\frac{\epsilon\sqrt{n}}{\sigma_X}) - 1 \\ &= 2\Phi(\lim_{n\to\infty} \frac{\epsilon\sqrt{n}}{\sigma_X}) - 1 \\ &= 2 - 1 = 1\end{aligned}$$

1 P

   (b) $\lim_{n\to\infty} P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) = 1 \Rightarrow$ konsistent 1 P

4. S_n^2 =1.3129 1 P

5. (a) $T_n = (n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1) \rightarrow t_7 = 7.8$ 1.5 P

   (b) $\chi^2_{0.025;6} = 1,24$ und $\chi^2_{0.975;6} = 14,45$ 1 P

   (c) $t_7 \not< 1,24$ und $t_7 \not> 14,45 \Rightarrow H_0$ kann bei einer Irrtumswahrscheinlichkeit von 5% nicht abgelehnt werden 1.5 P

   (d) (c) 0.5 P

6. $P(41,8 \leq X \leq 42,5) = 0.2708$ 1.5 P

Lösung 2

1. a) qualitativ 0.5 P

   b) 15 0.5 P

   c) $\frac{7}{15} = 0.4667$ 0.5 P

   d) $\frac{2}{15} = 13.33\%$ 0.5 P

2. a) 23 0.5 P

   b) 31 0.5 P

   c) linksschief, 0.5 P
   da Median näher am oberen Quantil (oder Modus > Median) 0.5 P

   d) -1.0415 0.5 P

   e) $\nu = \frac{\sigma}{\mu} = 0.2845$ 0.5 P

   f) $n \geq \left(\lambda_{1-\alpha/2}\frac{\sigma}{\epsilon}\right)^2 = \left(2.5758\frac{\sqrt{59}}{5}\right)^2$ 0.5 P
   $= 15.658$ Man müsste also 16 Inserate betrachten. 0.5 P

3. a) Fehler 1. Art: Nullhypothese ablehnen, obwohl sie richtig ist 0.5 P

   b) $T_n = \sqrt{n}\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{27}\frac{0.2-0.3}{\sqrt{0.3*0.7}} = -1.1339$ 0.5 P + 0.5 P

   c) $-\lambda_{0.9} = -1.2815$ 0.5 P

   d) $-1.1339 > -1.2815$, 0.5 P
   d.h. die Nullhypothese kann bei $\alpha = 0.10$ nicht abgelehnt werden. 0.5 P

   e) $\kappa = (1 - 6p(1-p))/(p(1-p)) = (1 - 6 * 0.3(0.7))/0.3 * 0.7 = -1.2381$ 0.5 P

   f) platykurtische Verteilung 0.5 P

Lösung 3

1) Absolut skaliert 0.5 P

2) a) $P(X = 2) = f_{Bin}(2; 10, 0.2) = 0.3020$ 0.5 P

   b) $P(X \leq 3) = F_{Bin(10,0.2)}(3) = 0.8791$ 0.5 P

   c) $P(10 - X \geq 4) = P(X \leq 6) = F_{Bin(10,0.2)}(6) = 0.9991$ 0.5 P

   d) $P(1 < X < 5) = F_{Bin(10,0.2)}(4) - F_{Bin(10,0.2)}(1) = 0.9672 - 0.3758 = 0.5914$ 1 P

3) a) $E[X] = np = 10 \cdot 0.2 = 2$ $0.5P$

   b) $VAR[X] = np(1 - p) = 2 \cdot 0.8 = 1.6$ $0.5P$

   c) $m_X = \lfloor np + p \rfloor = \lfloor 2 + 0.2 \rfloor = \lfloor 2.2 \rfloor = 2$ 0.5 P

4) a) $E[Y] = E[a + bX] = a + bE[X] = a + 1.2b$ $0.5P$

   b) $VAR[Y] = VAR[a + bX] = b^2 VAR[X] = 2b^2$ $0.5P$

5) a) $K \sim N(3\mu_W, 9\sigma_W^2)$ $0.5P$

   b) $J \sim N(\mu_z, \sigma_Z^2 + 1)$ $1P$

   c) $H \sim \chi^2(2)$ $1P$

5) a) $f(x) \geq 0 \ \forall \ x$ und $\int_{-\infty}^{\infty} f(x)dx = 1$ $0.5P$

   b)

$$\int_{-1}^{2} f(x)dx = \int_{-1}^{0} \alpha(1 + x^2)dx + \int_{0}^{2} \alpha x dx = 1$$

$$[\alpha x + \frac{\alpha x^3}{3}]_{-1}^{0} + [\frac{\alpha x^2}{2}]_{0}^{2} = 0 + 0 - (-\alpha - \frac{\alpha}{3}) + 2\alpha = 3\frac{1}{3}\alpha = 1$$

$$\Rightarrow \quad \alpha = 0.3$$ $1.5P$

Lösung 4

1.

(a) $P(Y = 1) = 0.2707$ (0.5P)

(b) $P(Y \leq 6) = 0.9955$ (0.5P)

(c) $P(Y \geq 2) = 0.5940$ (1P)

2. $P(Y \leq y) = 0.9989$
   $y = 7$ (1P)

3. großes n, kleines p mit $\lambda = np$ (1P)

4. $(2 - 4\sqrt{2}; 2 + 4\sqrt{2})$ (1P)

5. $P(\mu - 4\sigma < Y < \mu + 4\sigma) \geq \frac{15}{16}$ (0.5P)

6. $F(3) - F(0) = 0.7218$ (1P)

7. $L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i}$
   $LL(\lambda) = \sum_{i=1}^{n} \log \lambda e^{-\lambda X_i} = n \log(\lambda) - \lambda \sum_{i=1}^{n} X_i$ (1.5P)

8. $\frac{d \log L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i \stackrel{!}{=} 0$
   $\lambda_{ML} = \frac{n}{\sum_{i=1}^{n} X_i}$ (1P)

9. $\hat{\lambda}_{ML} = \frac{5}{15.1} = 0.3311$ (0.5P)

9. Eindeutig, asymptotisch erwartungstreu. (0.5P)

Klausur Statistik (10 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Freitag, 11.08.2017
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

Sie waren am Wochenende mit Ihren Freunden Kartfahren und haben sieben Runden absolviert. Sie betrachten die Zufallsvariable X: „Zeit pro Runde (in Sekunden)". Gehen Sie davon aus, dass gilt $X \sim N(\mu_X, \sigma_X^2)$.

1. Wie groß ist die Wahrscheinlichkeit in der 8. Runde eine Zeit von genau $42,\overline{3}$ Sekunden zu erreichen?

2. Sie berechnen das Stichprobenmittel als $\bar{x}_7 = 42.2286$. Wie würde sich das Stichprobenmittel tendenziell verändern, wenn Sie aufgrund eines technischen Defekts für die dritte Runde 70 Sekunden statt 41.7 Sekunden gebraucht hätten? (Hinweis: Eine Berechnung ist nicht notwendig.)

3.(a) Sie haben für einen allgemeinen Stichprobenumfang n und beliebiges $\epsilon > 0$ folgende Formel gegeben:

$$P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) = 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma_X}\right) - 1$$

Zeigen Sie mit Hilfe dieser Formel: $\lim_{n\to\infty} P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) = 1$.

(b) Ist das Stichprobenmittel $\bar{X}_n$ ein konsistenter Schätzer für μ_X? Begründen Sie Ihre Entscheidung.

4. Berechnen Sie mit Hilfe von $\sum_{i=1}^{7} x_i^2 = 12490.66$ und $\bar{x}_7 = 42.2286$ die Stichprobenvarianz.

5. Gehen Sie nun von einer Stichprobenvarianz von $s_7^2 = 1.3$ Sek^2 aus. Sie wollen bei einer Irrtumswahrscheinlichkeit von 5% folgendes Hypothesenpaar testen:

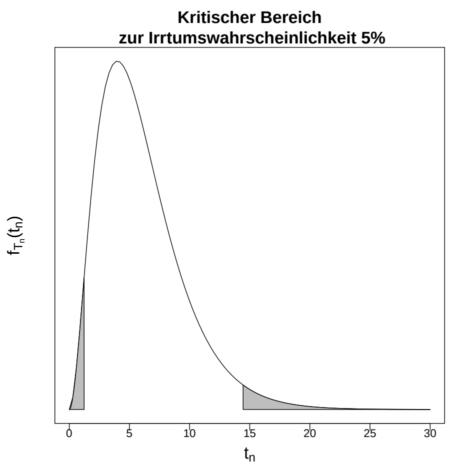
$$H_0 : \sigma^2 = 1 \text{ vs. } H_1 : \sigma^2 \neq 1$$

(a) Geben Sie die Prüfgröße sowie deren theoretische Verteilung an und berechnen Sie den Wert der realisierten Prüfgröße.

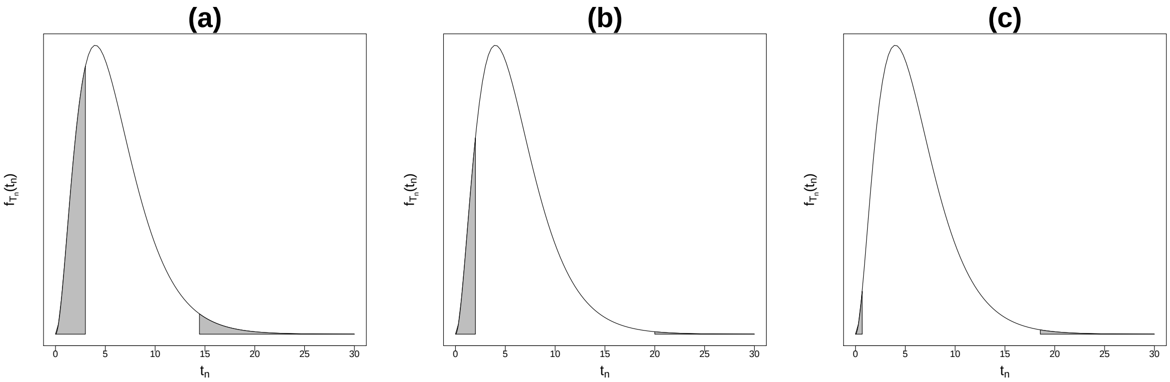
(b) Geben Sie die kritischen Schranken des Tests an.

(c) Treffen Sie eine Testentscheidung und formulieren Sie einen Antwortsatz. (Hinweis: Falls Sie in Teilaufgabe (b) bzw. (c) kein Ergebnis erhalten haben, benutzen Sie für die untere Schranke des kritischen Bereichs den Wert 3, für die obere Schranke des kritischen Bereichs den Wert 18 und für die realisierte Prüfgröße den Wert 4.)

(d) Im Folgenden bezeichne f die Dichtefunktion unter der Nullhypothese und $T_n = (n-1)S_n^2$. Grafisch lässt sich der kritische Bereich zu einer Irrtumswahrscheinlichkeit von 5% wie folgt darstellen:



Angenommen Sie hätten eine Irrtumswahrscheinlichkeit von 1% festgelegt. Welche der folgenden drei Grafiken würde sich dann ergeben?



Nehmen Sie nun an, die wahre Verteilung der Zufallsvariable ist gegeben durch:

$X \sim N(42, 1)$

6. Berechnen Sie die Wahrscheinlichkeit in der 8. Runde eine Zeit zwischen 41.8 und 42.5 Sekunden zu erreichen.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

Sie suchen eine Wohnung in Nürnberg in Nähe des Stadtzentrums.

1. Nach längerer Suche erhalten Sie folgende Aufstellung der Anzahl von Wohnungen gegliedert nach Nähe zum Stadtzentrum und in welchem Medium Sie das Inserat fanden:

Medium	Nähe Stadtzentrum	Nicht in Nähe Stadtzentrum
Internetportal	5	2
Schwarzes Brett der Universität	1	1
Zeitungsanzeige	2	4

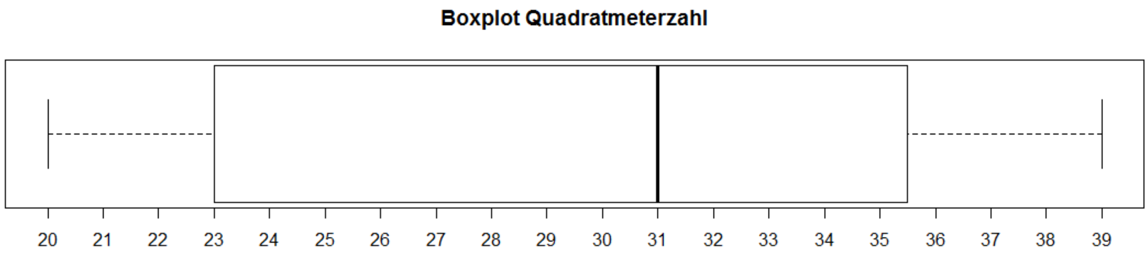
a) Welchen Merkmalstyp hat das Merkmal 'Medium'?

b) Wie viele Inserate haben Sie in der Tabelle gelistet?

c) Berechnen Sie den Anteil der Wohnungen, die nicht in Nähe des Stadtzentrums liegen.

d) Berechnen Sie den Anteil der Wohnungen, die in Stadtnähe liegen und die Sie als Zeitungsanzeige gefunden haben.

2. Sie listen nun die Wohnfläche [in qm] (X) der betrachteten Wohnungen auf und visualisieren die Daten mit folgendem Boxplot:



Der Modus liegt bei $x = 35$.

a) Geben Sie das untere Quartil ($x_{0.25}$) der Verteilung der Wohnfläche an.

b) Geben Sie den Median (x_{med}) der Verteilung der Wohnfläche an.

c) Ist die unimodale Verteilung der Wohnfläche rechtsschief, linksschief oder symmetrisch? Begründen Sie Ihre Antwort.

Nehmen Sie nun und im Folgenden an, die Verteilung der Wohnfläche folge einer Normalverteilung mit Mittelwert $\mu = 27$ und Varianz $\sigma^2 = 59$.

d) Berechnen Sie den ersten Pearsonschen Schiefekoeffizienten.

e) Geben Sie den Variationskoeffizienten an.

f) Wie viele Inserate müssen Sie betrachten, um mit einer Genauigkeit von $\pm$ 5 Quadratmetern und einer Irrtumswahrscheinlichkeit von 1% sagen zu können, wie groß die Wohnung, die sich sich leisten können, im Mittel sein wird?

Sei B: ′Besichtigte Wohnung besitzt einen Balkon′.

3. Von 27 Wohnungen die Sie besichtigen, haben 20% einen Balkon. Testen Sie Ihre Vermutung, dass die Wahrscheinlichkeit, dass eine Wohnung in Ihrer Preisklasse einen Balkon besitzt unter 30% liegt. Führen Sie einen Hypothesentest mit einer Irrtumswahrscheinlichkeit von 10% durch, wobei $B \sim Bernoulli(.)$.

a) Beschreiben Sie verbal den Fehler 1. Art.

b) Berechnen Sie die Prüfgröße des approximativen Tests.

c) Geben Sie die kritische Schranke für den Test an.

Hinweis: p-Wert= 0.1284.

d) Treffen Sie eine Testentscheidung und begründen Sie diese.

Nehmen Sie an, die Wahrscheinlichkeit, dass eine Wohnung einen Balkon hat, liegt bei 30%.

e) Berechnen Sie die Kurtosis der Verteilung von B.

Sollten Sie keine Lösung für Aufgabe 3.e) bekommen haben, verwenden Sie: $\kappa = -1.5$.

f) Weist der Wert der Kurtosis auf eine mesokurtische, leptokurtische oder platykurtische Verteilung hin?

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

Sie sind Qualitätsprüfer bei der Firma „Gear“, welche Getriebe produziert. Ihre Aufgabe ist es zu prüfen, ob die produzierten Getriebe defekt sind. Nehmen Sie an, die Zufallsvariable

X : „Anzahl der defekten Getriebe”

ist binomialverteilt mit den Parametern $n = 10$ und $p = 0.2$.

1. Nennen Sie das höchste Skalenniveau von X.

2. Berechnen Sie die Wahrscheinlichkeit, dass

   (a) genau zwei Getriebe defekt sind.

   (b) höchstens drei Getriebe defekt sind.

   (c) mindestens vier Getriebe nicht defekt sind.

   (d) mehr als ein Getriebe, aber weniger als fünf defekt sind.

3. Bestimmen Sie,

   (a) den Erwartungswert von X.

   (b) die Varianz von X.

   (c) den Modus von X.

4. Gehen Sie unabhängig von den vorherigen Ergebnissen davon aus, dass gilt:

   $E[X] = 1.2$ und $VAR[X] = 2$.

   Gegeben sei die Zufallsvariable $Y = a + bX$, mit $a, b \in \mathbb{R}$. Berechnen Sie

   (a) den Erwartungswert von Y.

   (b) die Varianz von Y.

Betrachten Sie die folgenden vier voneinander unabhängigen Zufallsvariablen:

$$W \sim N(\mu_W, \sigma_W^2),$$
$$Z \sim N(\mu_Z, \sigma_Z^2),$$
$$A \sim N(0, 1),$$
$$B \sim N(0, 1).$$

5. Geben Sie die Namen der Verteilung, **inklusive** aller Parameter, der nachfolgenden Zufallsvariablen an:

   (a) $K = 3W$

   (b) $J = Z + A$

   (c) $H = A^2 + B^2$

6. Gegeben sei folgende nicht-negative Funktion, mit $\alpha > 0$:

$$f_X(x) = \begin{cases} \alpha(1+x^2) & \text{für} \quad -1 \leq x \leq 0, \\ \alpha x & \text{für} \quad 0 \leq x \leq 2, \\ 0 & \text{sonst.} \end{cases}$$

(a) Nennen Sie die zwei Eigenschaften die eine Dichtefunktion besitzen muss.

(b) Welchen Wert muss α annehmen, damit f_X als Dichtefunktion für eine Zufallsvariable X interpretiert werden kann?

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

In Ihrem Workspace liegt ein Dataframe `airports`, welcher Informationen zu 6344 Flughäfen weltweit enthält[1]:

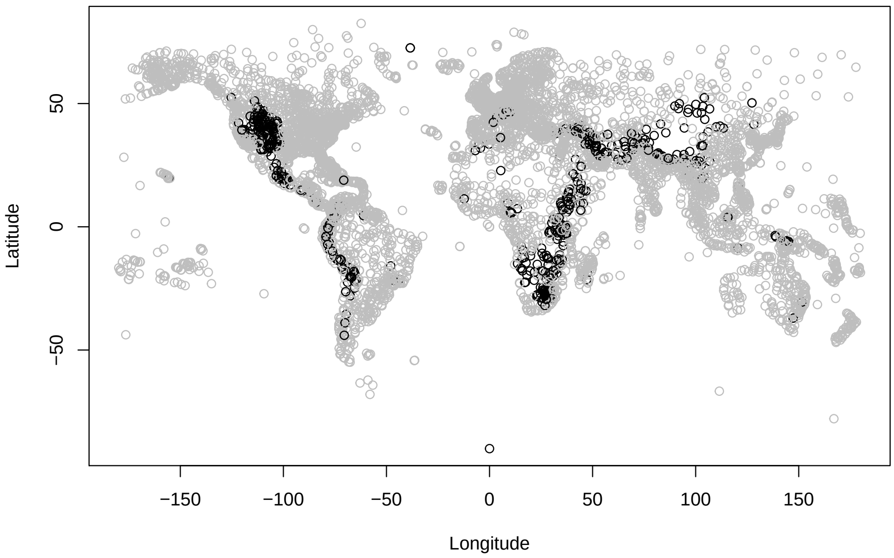
Spaltenname	**Info**
`AirportID`	Identifikationsnummer
`Name`	Name des Flughafens
`City`	Name der Stadt nahe des Flughafens
`Country`	Land
`IATA`	Flughafencode-Kürzel 1
`ICAO`	Flughafencode-Kürzel 2
`Latitude`	Position des Flughafens (Längengrad)
`Longitude`	Position des Flughafens (Breitengrad)
`Altitude`	Höhe des Flughafens über dem Meeresspiegel [m]
`Timezone`	Zeitdifferenz zur koordinierten Weltzeit (UTC)

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

```
airports[1:4,-c(1,2,5,6)]
wahrfalsch = airports$Altitude > 1000
farbe = ifelse(wahrfalsch,"black","grey")
plot(airports$Longitude,airports$Latitude,col = farbe)
```

```
         City          Country  Latitude Longitude  Altitude Timezone
1      Goroka Papua New Guinea  -6.0817  145.3919 1609.9535       10
2      Madang Papua New Guinea  -5.2071  145.7887    6.0960       10
3 Mount Hagen Papua New Guinea  -5.8268  144.2959 1642.2623       10
4      Nadzab Papua New Guinea  -6.5698  146.7262   72.8472       10
```

[1]http://www.openflights.org/data.html



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

1. Welche Dimensionen hat der Dataframe `airports`?

2. Was bedeuten die schwarzen Punkte auf der Landkarte?

3. Geben Sie den `R`-Code für die Berechnung der größten Höhe über dem Meeresspiegel der beobachteten Flughäfen an.

4. Welchen Output liefert der Befehl `airports$Longitude[seq(1,3,2)]`?

5. Geben Sie den `R`-Code für die Berechnung der durchschnittlichen Höhe über dem Meeresspiegel an, bedingt darauf, dass ein Flughafen höher als 2000 Meter über dem Meeresspiegel gelegen ist.

6. Mit welchem `R`-Code können Sie sich den Teildatensatz `teil1` abspeichern, der lediglich Flughäfen enthält, welche eine Zeitdifferenz von 0 Stunden zur koordinierten Weltzeit haben?

7. Mit welchem `R`-Code können Sie sich den Teildatensatz `teil2` abspeichern, der nur aus den Spalten `Latitude`, `Longitude` und `Altitude` besteht?

8. Mit welchem R-Code können Sie den Modus des Merkmals Timezone berechnen? Hinweis: Der Modus ist hier eindeutig.

In Ihrem Workspace finden Sie nun folgende Funktion:

```
f = function(y,x,r){
  (-1+x+y^2)^(1/r)+1
}
```

9. Geben Sie den R-Code an, mit dem Sie das Minimum der obigen Funktion für die Werte $x = 2$ und $r = 4$ numerisch bestimmen können. Der Startwert beträgt 2.

10. Vervollständigen Sie den folgenden R-Code, um eine Funktion zu erhalten, mit der man die Likelihood-Funktion der Chi-Quadrat-Verteilung mit Freiheitsgrad `nu` gegeben dem Stichprobenvektor `stich` numerisch maximieren kann:

```
L_chisq = function(       ,       ){
    prod(       (       ,       )       )
}
```

Schmierpapier zu Aufgabe 4

Lösung Klausur SoSe 2017 (10 ECTS)

Lösung 1

1. $P(X = 42,\overline{3}) = 0$ 0.5 P

2. Steigt 0.5 P

3. (a)

$$\begin{aligned}\lim_{n\to\infty} P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) &= \lim_{n\to\infty} 2\Phi(\frac{\epsilon\sqrt{n}}{\sigma_X}) - 1 \\ &= 2\Phi(\lim_{n\to\infty} \frac{\epsilon\sqrt{n}}{\sigma_X}) - 1 \\ &= 2 - 1 = 1\end{aligned}$$

1 P

   (b) $\lim_{n\to\infty} P(\mu_X - \epsilon \leq \bar{X}_n \leq \mu_X + \epsilon) = 1 \Rightarrow$ konsistent 1 P

4. S_n^2 =1.3129 1 P

5. (a) $T_n = (n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1) \to t_7 = 7.8$ 1.5 P

   (b) $\chi^2_{0.025;6} = 1,24$ und $\chi^2_{0.975;6} = 14,45$ 1 P

   (c) $t_7 \not< 1,24$ und $t_7 \not> 14,45 \Rightarrow H_0$ kann bei einer Irrtumswahrscheinlichkeit von 5% nicht abgelehnt werden 1.5 P

   (d) (c) 0.5 P

6. $P(41,8 \leq X \leq 42,5) = 0.2708$ 1.5 P

Lösung 2

1. a) qualitativ — 0.5 P

   b) 15 — 0.5 P

   c) $\frac{7}{15} = 0.4667$ — 0.5 P

   d) $\frac{2}{15} = 13.33\%$ — 0.5 P

2. a) 23 — 0.5 P

   b) 31 — 0.5 P

   c) linksschief, — 0.5 P
   da Median näher am oberen Quantil (oder Modus > Median) — 0.5 P

   d) -1.0415 — 0.5 P

   e) $\nu = \frac{\sigma}{\mu} = 0.2845$ — 0.5 P

   f) $n \geq \left(\lambda_{1-\alpha/2}\frac{\sigma}{\epsilon}\right)^2 = \left(2.5758\frac{\sqrt{59}}{5}\right)^2$ — 0.5 P
   $= 15.658$ Man müsste also 16 Inserate betrachten. — 0.5 P

3. a) Fehler 1. Art: Nullhypothese ablehnen, obwohl sie richtig ist — 0.5 P

   b) $T_n = \sqrt{n}\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{27}\frac{0.2-0.3}{\sqrt{0.3*0.7}} = -1.1339$ — 0.5 P + 0.5 P

   c) $-\lambda_{0.9} = -1.2815$ — 0.5 P

   d) $-1.1339 > -1.2815$, — 0.5 P
   d.h. die Nullhypothese kann bei $\alpha = 0.10$ nicht abgelehnt werden. — 0.5 P

   e) $\kappa = (1 - 6p(1-p))/(p(1-p)) = (1 - 6 * 0.3(0.7))/0.3 * 0.7 = -1.2381$ — 0.5 P

   f) platykurtische Verteilung — 0.5 P

Lösung 3

1) Absolut skaliert 0.5 P

2)
   a) $P(X = 2) = f_{Bin}(2; 10, 0.2) = 0.3020$ 0.5 P
   b) $P(X \leq 3) = F_{Bin(10,0.2)}(3) = 0.8791$ 0.5 P
   c) $P(10 - X \geq 4) = P(X \leq 6) = F_{Bin(10,0.2)}(6) = 0.9991$ 0.5 P
   d) $P(1 < X < 5) = F_{Bin(10,0.2)}(4) - F_{Bin(10,0.2)}(1) = 0.9672 - 0.3758 = 0.5914$ 1 P

3)
   a) $E[X] = np = 10 \cdot 0.2 = 2$ $0.5P$
   b) $VAR[X] = np(1 - p) = 2 \cdot 0.8 = 1.6$ $0.5P$
   c) $m_X = \lfloor np + p \rfloor = \lfloor 2 + 0.2 \rfloor = \lfloor 2.2 \rfloor = 2$ 0.5 P

4)
   a) $E[Y] = E[a + bX] = a + bE[X] = a + 1.2b$ $0.5P$
   b) $VAR[Y] = VAR[a + bX] = b^2 VAR[X] = 2b^2$ $0.5P$

5)
   a) $K \sim N(3\mu_W, 9\sigma_W^2)$ $0.5P$
   b) $J \sim N(\mu_z, \sigma_Z^2 + 1)$ $1P$
   c) $H \sim \chi^2(2)$ $1P$

5)
   a) $f(x) \geq 0 \; \forall \; x$ und $\int_{-\infty}^{\infty} f(x)dx = 1$ $0.5P$
   b)

$$\int_{-1}^{2} f(x)dx = \int_{-1}^{0} \alpha(1 + x^2)dx + \int_{0}^{2} \alpha x dx = 1$$

$$[\alpha x + \frac{\alpha x^3}{3}]_{-1}^{0} + [\frac{\alpha x^2}{2}]_0^2 = 0 + 0 - (-\alpha - \frac{\alpha}{3}) + 2\alpha = 3\frac{1}{3}\alpha = 1$$

$$\Rightarrow \quad \alpha = 0.3$$ $1.5P$

Lösung 4

1. 6344, 10 — 1 P
2. Flughäfen, die mehr als 1000 m über dem Meeresspiegel liegen — 0.5 P
3. `max(airports$Altitude)` — 0.5 P
4. 145.3919 144.2959 — 0.5 P
5. `mean(airports$Altitude[airports$Altitude > 2000])` — 1.5 P
6. `teil1 = airports[airports$Timezone == 0,]` — 1.5 P
7. `teil2 = airports[,c(7,8,9)]` — 1.5 P
8. `which.max(table(airports$Timezone))` — 1.0 P
9. `nlm(f,p=2,x = 2,r = 4)` — 1.0 P
10.
```
L_chisq = function(stich,nu){
    prod(dchisq(stich,nu))
}
```
   1.0 P