Klausur Statistik (7.5 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Dienstag, 31.07.2018
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 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 prinzipiell 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

1. Sie interessieren sich für das Verhalten der Fernsehzuschauer während der Fußball WM. Dabei möchten Sie wissen ob das Ansehen des WM Finales unabhängig vom Geschlecht des Fernsehzuschauers ist. Dazu betrachten Sie die Eigenschaften:

   A : "Fernsehzuschauer ist weiblich"

   B : "Fernsehzuschauer hat das WM Finale geschaut"

   Folgende relativen Häufigkeiten haben Sie in einer Vierfeldertafel zusammengefasst:

	A	$\bar{A}$	
B	0.25	0.1	0.35
$\bar{B}$	**a)**	0.5	**b)**
	0.4	**c)**	1

a) Welcher Wert fehlt in der Zelle a)?

b) Welcher Wert fehlt in der Zelle b)?

c) Welcher Wert fehlt in der Zelle c)?

d) Beschreiben Sie die Menge $A \cap B$ verbal im Kontext der Situation.

e) Beschreiben Sie die Menge $A \cup \bar{A}$ verbal im Kontext der Situation.

f) Welcher Anteil der Zuschauer, von denen man weiß, dass sie weiblich sind, hat das WM Finale angesehen?

g) Sind die Ereignisse A und B statistisch unabhängig (mit Begründung)?

2. Sie wissen, dass ein 40 prozentiger Anteil der Kinder das WM Finale angesehen hat. Wie groß ist die Wahrscheinlichkeit, dass von 20 unabhängig voneinander ausgewählten Kindern genau 13 das WM Finale angesehen haben?

3. Sie veranstalten Public Viewings während der Fußball WM und möchten Rückschlüsse auf den Getränkekonsum der Zuschauer ziehen. Dazu betrachten Sie folgende Zufallsvariable:

   X : „Verkaufte Getränkemenge in Liter während eines Public Viewings".

   Sie gehen davon aus, dass X normalverteilt ist und berechnen die Stichprobenvarianz von einer einfachen Stichprobe mit Stichprobenumfang $n = 5$ als $s_5^2 = 2.5$. Nun möchten Sie testen ob die Varianz kleiner als 2 ist und stellen dazu folgendes Hypothesenpaar auf:

$$H_0 : \sigma^2 \geq 2 \quad \text{vs.} \quad H_1 : \sigma^2 < 2.$$

a) Geben Sie die zugehörige Teststatistik und deren theoretische Verteilung an.

b) Berechnen Sie den Wert der Teststatistik.

c) Berechnen Sie den p-Wert. Falls Sie Teilaufgabe b) nicht lösen konnten, verwenden Sie den Wert $t_5 = 3$. Außerdem haben Sie die folgende Werte gegeben:

$$F_{\chi^2(5)}(5) = 0.5841,\ F_{\chi^2(4)}(5) = 0.7127, F_{\chi^2(4)}(4) = 0.5940,\ F_{\chi^2(5)}(4) = 0.4506,$$
$$F_{\chi^2(4)}(3) = 0.4422,\ F_{\chi^2(5)}(3) = 0.3000,\ F_{\chi^2(3)}(4) = 0.7385$$

d) Treffen Sie eine Testentscheidung mit einer Irrtumswahrscheinlichkeit von 10% (mit Begründung). Falls Sie Teilaufgabe c) nicht lösen konnten, verwenden Sie 0.5 als p-Wert.

4. Die Verteilung einer Zufallsvariable ist gegeben durch $X \sim \mathcal{N}(20, 4)$.

a) Berechnen Sie die Wahrscheinlichkeit des 3-fachen zentralen Schwankungsintervalls für X.

b) Geben Sie das 70% Quantil von X an.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

In einer Bank werden die einzelnen Kreditnehmer verschiedenen Ratingklassen zugeordnet, wobei AAA die beste und C die schlechteste Ratingkategorie beschreibt. Das Rating eines Kreditnehmers sei durch die Zufallsvariable X beschrieben:

Ratingklasse x	AAA	AA	A	BBB	BB	B	C
$P(X = x)$	0.08	0.12	0.28	0.18	0.11	0.08	θ

1. Welchen Wert muss θ annehmen, damit eine Wahrscheinlichkeitsverteilung vorliegt?

2. Welche Ausprägungen nehmen Median und Modus von X an?

3. Berechnen Sie die Gini Entropie für X. (Hinweis: Falls Sie Teilaufgabe 1 nicht lösen konnten, verwenden Sie $\theta = 0.1$)

Im Folgenden werden die Ratingklassen zweier Kreditnehmer X_1 und X_2 betrachtet, die beide jeweils der oben angegebenen Verteilung folgen.

4. Berechnen Sie die Wahrscheinlichkeit $P(X_1 = AAA|X_2 = AAA)$, dass Kreditnehmer X_1 Ratingklasse AAA besitzt, gegeben dass Kreditnehmer X_2 bereits in Kategorie AAA eingestuft wurde, wenn $P(X_1 = AAA \cap X_2 = AAA) = 0.05$ ist.

5. Wie groß ist die Wahrscheinlichkeit $P(X_1 = AAA|X_2 = AAA)$ unter der Voraussetzung, dass X_1 und X_2 stochastisch unabhängig sind?

Im Folgenden sind Realisationen der täglichen Rendite R_i einer Aktie von 10 aufeinander folgenden Tagen gegeben.

Tag i	1	2	3	4	5	6	7	8	9	10
Tägliche Rendite r_i	0.04	-0.04	0.04	0.02	-0.03	0.02	0.01	-0.01	-0.03	0.00

6. Geben Sie sowohl Merkmalstyp als auch maximales Skalenniveau des Merkmals "Tägliche Rendite" an.

7. Der Wert der Aktie zu Beginn des 10-Tages-Zeitraums beträgt 75 Euro, welchen Wert hat die Aktie zum Ende des 10-Tages-Zeitraums?

Nehmen Sie nun an, dass die Renditen R_i unabhängig und identisch normalverteilt sind mit Parametern μ_R und σ_R^2.

8. Berechnen Sie das Stichprobenmittel $\bar{R}_{10}$! Hinweis: $\sum_{i=1}^{10} r_i = 0.02$

9. Berechnen Sie die Stichprobenvarianz S_{10}^2! Hinweis: $\sum_{i=1}^{10} r_i^2 = 0.0076$

10. Nehmen Sie nun an, dass $\bar{r}_{10} = 0.001$ und $S_{10}^2 = 0.0006$ ist. Berechnen Sie das realisierte, symmetrische 95% Konfidenzintervall für den Mittelwert.

11. Gehen Sie nun davon aus, dass das realisierte 95% Konfidenzintervall für den Mittelwert durch $[-0.004, 0.007]$ gegeben ist. Begründen Sie kurz, ob Sie die zugehörige Hypothese H_0: $\mu_R = 0$ verwerfen können oder nicht.

12. Nehmen Sie nun an, dass die Momente des Stichprobenmittels $\bar{R}$ bekannt sind, wobei $\mu_{\bar{R}} = 0$ und $\sigma^2_{\bar{R}} = 0.0009$. Bestimmen Sie die Grenzen eines zentralen Schwankungsintervalls in dem $\bar{R}$ mit 95% Wahrscheinlichkeit enthalten ist mit Hilfe der Chebyshevschen Ungleichung.

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

Sie interessieren sich für die wirtschaftliche Leistung der Kokosinseln im Indischen Ozean. Um diese näher zu analysieren, definieren Sie die Zufallsvariable X, welche einer stetigen Verteilung folgt:

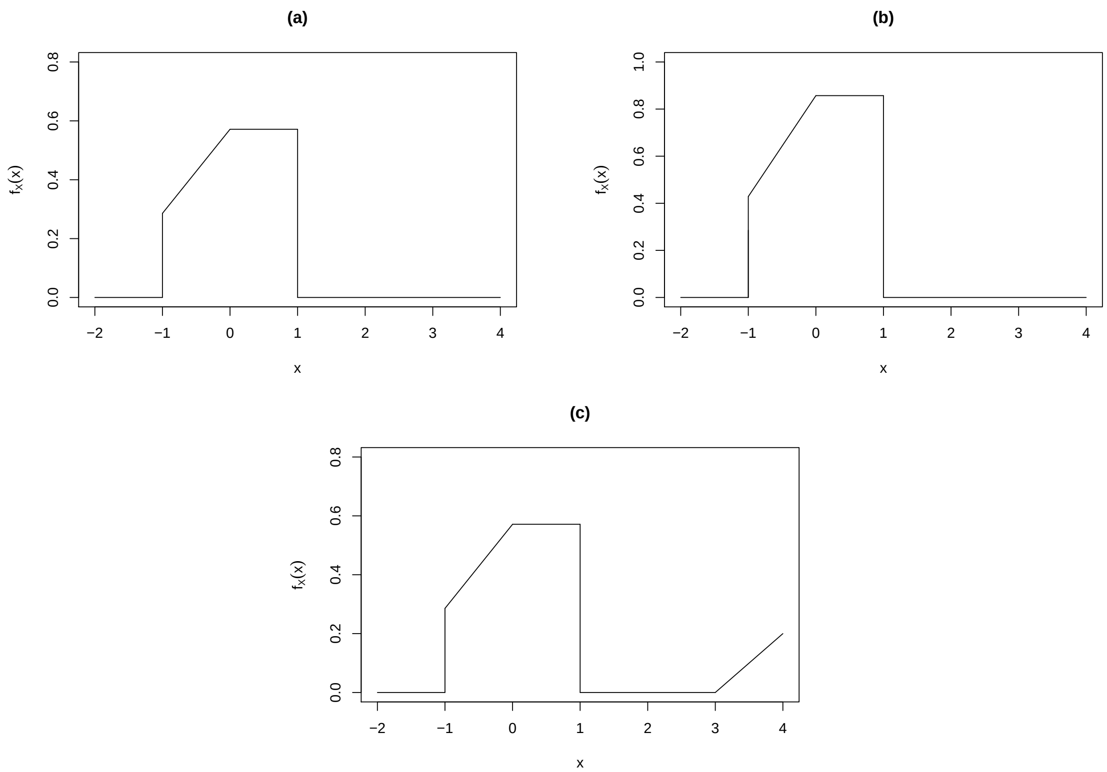
X : „Relatives Wachstum des Bruttoinlandsprodukts (BIP) auf den Kokosinseln" .

1. Wie groß ist die Wahrscheinlichkeit, dass das Wachstum des BIP genau 0.1 beträgt?

Gegeben sei außerdem folgende Funktion mit $c > 0$:

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

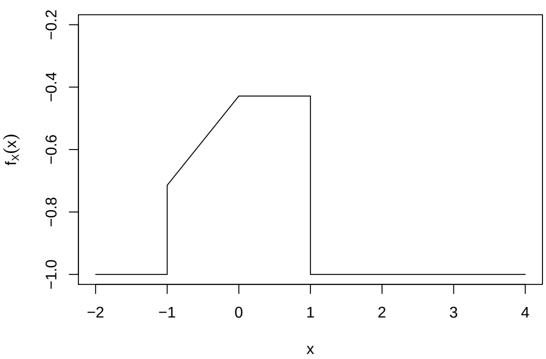
2. Bei welcher der folgenden Abbildungen kann es sich **nicht** um die oben genannte Funktion handeln?



3. Wie muss c gewählt werden, damit es sich bei $f_X(x)$ um eine Dichte handelt? Es ist eine Berechnung erforderlich.

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

4. Warum kann es sich bei folgender Abbildung **nicht** um eine Dichte handeln?



5. Bestimmen Sie das zweite Moment $E[X^2]$.
   Verwenden Sie hierbei: $E[X] = 0$ und $Var[X] = 0.5$.

Sei nun G eine Poisson-verteilte Zufallsvariable mit $G \sim Pois(\lambda), \lambda = 3$.

6. Welcher der folgenden 3 Träger gehört zur diskreten Zufallsvariable G?
   (a) $D(G) = \mathbb{R}$
   (b) $D(G) = [a, b], a, b \in \mathbb{R}, a < b$
   (c) $D(G) = \{0, 1, 2, ...\}$

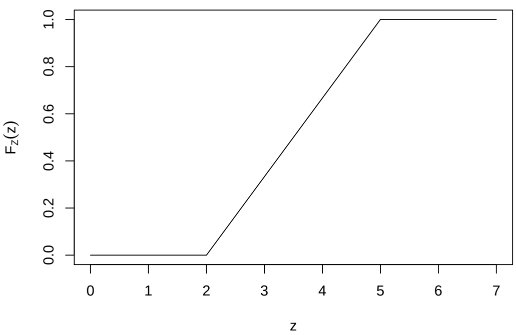
7. Berechnen Sie die Wahrscheinlichkeit, dass G größer als 2, aber kleiner als 5 ist.

8. Die Wahrscheinlichkeit, dass G eine Schwelle g überschreitet, beträgt 0.5768. Bestimmen Sie g.

In Abhängigkeit von G ($G \sim Pois(\lambda), \lambda = 3$) definieren Sie die Zufallsvariable $W = c(a + bG)$ mit $a, b, c \in \mathbb{R}$.

9. Berechnen Sie die Varianz von W.

Abgebildet sehen Sie die Verteilungsfunktion einer stetigen Zufallsvariable Z.

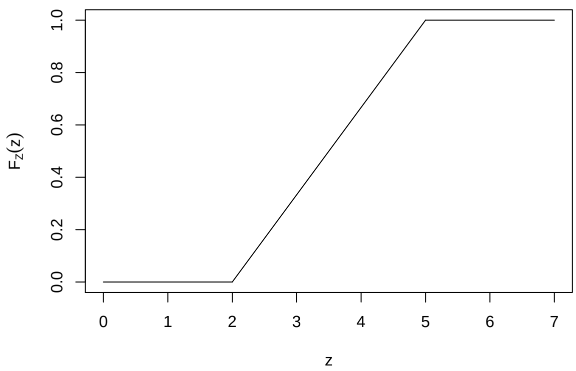


10. Welcher Verteilung folgt Z?

11. Bestimmen Sie mit Hilfe der Grafik $F_Z(5)$.

12. Bestimmen Sie mit Hilfe der Grafik $P(Z > 2)$.

13. Zeichnen Sie in folgender Grafik das 40%-Quantil ein und geben Sie dessen Wert an.



Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

Sie besitzen einen Kiosk.

1. Die Variable P: 'Anzahl Personen m, die in 5 Minuten Ihren Kiosk betreten' stamme aus einer Poissonverteilung, wobei die Annahmen an einen Poissonprozess erfüllt sind. Sie stoppen nun die Zeit in Minuten zwischen dem Ankommen zweier Personen (Z).

a) Welcher Verteilungsfamilie folgt die Variable Z (ohne Parameter)?

b) Beschreiben Sie allgemein die Methode der Momente.

c) Schätzen Sie den Parameter der Verteilung von Z mit Hilfe von $\overline{z}_m = 0.4$.

d) Berechnen Sie die Varianz der Verteilung von Z und verwenden Sie dafür $\lambda = 2.2$.

2. Jede Woche bestellen Sie von zwei unterschiedlichen Anbietern (Anbieter A und Anbieter B) Bonbons für Ihren Kiosk. Sie haben die Bonbons in den Geschmacksrichtungen süß und sauer bestellt. Sie wissen, dass die Wahrscheinlichkeit für ein saures Bonbon, wenn das Bonbon von Anbieter A kommt 30% und wenn das Bonbon von Anbieter B kommt 50% ist.

a) Welchen Merkmalstyp hat das Merkmal G: 'Geschmacksrichtung eines Bonbons'?

b) Wie ist das Merkmal G: 'Geschmacksrichtung eines Bonbons' skaliert?

c) Wenn Sie gleich viele Bonbons von Anbieter A und Anbieter B in einer Schüssel zusammenschütten, wie hoch ist dann die Wahrscheinlichkeit ein saures Bonbon daraus zu ziehen?

d) Wieviele Bonbons müssen Sie mindestens aus einer der Packungen ziehen, dass mit einer Irrtumswahrscheinlichkeit von 5% der Anteil der sauren Bonbons in Ihrer Stichprobe zur Grundgesamtheit um weniger als 10% vom wahren Anteil nach oben und unten abweicht? Nutzen Sie die Ungleichung von Chebyshev.

3. Nun haben Sie die Vermutung, dass ein Zusammenhang zwischen der Dauer in Minuten, die Ihre Kunden in Ihrem Kiosk verbringen (M) und dem Betrag in Euro, den Ihre Kunden bei Ihnen ausgeben (B) besteht. Sie haben eine einfache Stichprobe vom Umfang $n = 25$ gezogen und notieren sich die jeweiligen Realisationen von M und B.

   Sie vermuten, dass M und B korreliert sind.

a) Was bedeutet Korrelation?

Testen Sie Ihre Vermutung bei einer Irrtumswahrscheinlichkeit von 10%, wenn Sie wissen, dass (M,B) bivariat normalverteilt ist mit unbekanntem Erwartungswertvektor (μ_M, μ_B) und unbekanntem Varianzvektor (σ_M^2, σ_B^2).
Sie stellen folgende Hypothesen auf:

$$H_0 : \rho = 0 \qquad H_1 : \rho \neq 0$$

Sie haben außerdem gegeben, dass der Wert des Stichprobenkorrelationskoeffizienten $R = 0.7$ ist. Gehen Sie davon aus, dass die Approximationsbedingungen erfüllt sind.

b) Berechnen Sie den Wert der Teststatistik.

c) Wie lautet der Ablehnungsbereich mit konkretem Wert?

d) Nehmen Sie nun an, dass der Hypothesentest einen p-Wert von 0.01 liefert. Treffen Sie eine Testentscheidung auf dem 10% Signifikanzniveau.

Nun wollen Sie das approximative Konfidenzintervall für den Korrelationskoeffizienten ρ bilden.

e) Berechnen Sie hierfür zunächst Fishers Z-Transformation.

Verwenden Sie ab jetzt: $Z(R) = 0.8$

f) Berechnen Sie nun die untere Grenze des approximativen realisierten 90%-Konfidenzintervalls für ρ.

Schmierpapier zu Aufgabe 4

Lösung Klausur SoSe 18 (7.5 ECTS)

Lösung 1

1a)	0.15	0.5 P
1b)	0.65	0.5 P
1c)	0.6	0.5 P
1d)	Menge der Personen, die weiblich sind und das WM Finale gesehen haben.	0.5 P
1e)	Menge aller Fernsehzuschauer.	0.5 P
1f)	$P(B\|A) = 0.625$	1 P
1g)	$P(A \cap B) = 0.25 \neq P(A)P(B) = 0.14 \Rightarrow$ Nicht statistisch unabhängig.	1 P
2.	0.0146	0.5 P
3a)	$(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$	1 P
3b)	5	0.5 P
3c)	$p = 0.7127$	1 P
3d)	$p > \alpha \Rightarrow$ nicht ablehnen.	1 P
4a)	$P(-3 < \frac{X-\mu}{\sigma} < 3) = 0.9974$	1 P
4b)	21.0488	0.5 P

Lösung 2

1. $\theta = 0.15$ 0.5 P

2. $\bar{x}_{\text{mod}} = $ "A", $\bar{x}_{\text{med}} = $ "BBB" 1 P

3. $0.08 \cdot 0.92 + 0.12 \cdot 0.88 + \ldots + 0.15 \cdot 0.85 = 0.8274$
   Mit Ersatzergebnis: $\sum_i p_i(1-p_i) = 0.7899$, $1 - \sum_i p_i^2 = 0.8399$ 1 P

4. $P(X_1 = AAA | X_2 = AAA) = \frac{P(X_1 = AAA \cap X_2 = AAA)}{P(X_2 = AAA)} = \frac{0.05}{0.08} = 0.625$ 1 P

5. $P(X_1 = AAA | X_2 = AAA) = P(X_1 = AAA) = 0.08$ 0.5 P

6. Merkmalstyp: quantitativ, Skalenniveau: verhältnis-/absolut skaliert 1 P

7. $75[EUR] \cdot \prod_{i=1}^{10}(1 + r_i) = 76.23[EUR]$ 0.5 P

8. $\bar{R}_{10} = \frac{\sum_{i=1}^{10} r_i}{10} = 0.002$ 0.5 P

9. $S_{10}^2 = \frac{1}{9}\left(\sum_{i=1}^{10} r_i^2 - 10\bar{r}^2\right) = \frac{1}{9}\left(\sum_{i=1}^{10} r_i^2 - \frac{1}{10}\left(\sum_{i=1}^{10} r_i\right)^2\right)$
   $= 0.0008$ 1 P

10. $\left[\bar{r}_{10} - t_{1-\frac{\alpha}{2};9}\frac{S_{10}}{\sqrt{10}}; \bar{r}_{10} + t_{1-\frac{\alpha}{2};9}\frac{S_{10}}{\sqrt{10}}\right] = [-0.0165; 0.0185]$
    wobei $t_{1-\frac{\alpha}{2};9} = 2.262$ 1.5 P

11. H_0 beibehalten, da 0 innerhalb der Intervallgrenzen $[-0.004; 0.007]$. 0.5 P

12. $P(\mu_R - k \cdot \sigma_R < R < \mu_R + k \cdot \sigma_R) \geq 1 - \frac{1}{k^2} = 0.95 \Rightarrow k = \sqrt{20}$
    $\Rightarrow P(-0.1342 < R < 0.1342) = 0.95 \Rightarrow g_u = -0.1342, g_o = 0.1342$ 1 P

Lösung 3

1. 0 — 0.5 P
2. (c) — 0.5 P
3. $\int_{-1}^{0} c(2+x)dx + \int_{0}^{1} 2cdx = [2cx + \frac{c}{2}x^2]_{-1}^{0} + [2cx]_0^1 =$
   $-(-2c + \frac{c}{2}) + 2c = 2c - \frac{c}{2} + 2c = 3.5c = 1 \leftrightarrow c = \frac{2}{7}$ — 1.5 P
4. negative Funktionswerte / $f_X(x) \geq 0$ verletzt — 0.5 P
5. $Var(X) = E[X^2] - E[X]^2 \rightarrow E[X^2] = Var(X) + E[X]^2 = 0.5$ — 0.5 P
6. (c) — 0.5 P
7. $P(2 < G < 5) = P(G \leq 4) - P(G \leq 2) = 0.8153 - 0.4232 = 0.3921$
   (mit Wkeitsfkt 0.392) — 1 P
8. $P(G > g) = 0.5768 \rightarrow 1 - P(G \leq g) = 0.5768 \rightarrow P(G \leq g) = 0.4232$
   $\rightarrow g = 2$ — 1.5 P
9. $Var[c(a + bG)] = c^2b^2Var[G] = c^2b^2 3$ — 1 P
10. (stetige) Gleichverteilung / Rechteckverteilung — 0.5 P
11. 1 — 0.5 P
12. $P(Z > 2) = 1 - P(Z \leq 2) = 1 - 0 = 1$ — 1 P
13. 3 — 0.5 P

Lösung 4

1. a) Die Zeit zwischen dem Eintreten zweier Poisson-Ereignisse ist exponentialverteilt.

   0.5 P

   b) Methode der Momente: Potenzmomente der Stichprobe werden mit entsprechenden Potenzmomenten der Grundgesamtheit gleichgesetzt. 0.5 P

   c) $\hat{\lambda} = 1/\overline{Z}_m = 2.5$ 0.5 P

   d) Varianz: $1/\lambda^2 = 0.2066$. 0.5 P

2. a) Merkmalstyp: qualitativ 0.5 P

   b) Skalierung: Nominal 0.5 P

   c) 0.4 0.5 P

   d) $1/(4 * 0.05 * 0.1^2) = 500$ 1 P

3. a) Korrelation: (linearer) Zusammenhang zweier Zufallsvariablen 0.5 P

   b) Teststatistik: 4.7009 1 P

   c) Ablehnungsbereich: $|t_n| > t_{1-\alpha/2;n-2}$, wobei $t(0.95; 23) = 1.714$ 1 P

   d) Testentscheidung: Da $p = 0.01 < \alpha = 0.1$ ($|4.7009| > 1.714$) 0.5 P
   Die Nullhypothese kann auf dem 10% Signifikanzniveau abgelehnt werden.

   0.5 P

   e) Fishers Z-Transformation: Z(R)=0.8673 0.5 P

   f) $[\frac{e^{2(Z(R)-\lambda_{1-\alpha/2}\frac{1}{\sqrt{n-3}})}-1}{e^{2(Z(R)-\lambda_{1-\alpha/2}\frac{1}{\sqrt{n-3}})}+1}; \quad \cdot \quad]$ 0.5 P

   $\lambda_{1-\alpha/2} = 1.6448$ 0.5 P
   $[0.4213; \quad \cdot \quad]$ 0.5 P

Klausur Statistik (10 ECTS)

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Dienstag, 31.07.2018
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 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 prinzipiell 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

1. Sie interessieren sich für das Verhalten der Fernsehzuschauer während der Fußball WM. Dabei möchten Sie wissen ob das Ansehen des WM Finales unabhängig vom Geschlecht des Fernsehzuschauers ist. Dazu betrachten Sie die Eigenschaften:

   A : "Fernsehzuschauer ist weiblich"

   B : "Fernsehzuschauer hat das WM Finale geschaut"

   Folgende relativen Häufigkeiten haben Sie in einer Vierfeldertafel zusammengefasst:

	A	$\bar{A}$	
B	0.25	0.1	0.35
$\bar{B}$	**a)**	0.5	**b)**
	0.4	**c)**	1

a) Welcher Wert fehlt in der Zelle a)?

b) Welcher Wert fehlt in der Zelle b)?

c) Welcher Wert fehlt in der Zelle c)?

d) Beschreiben Sie die Menge $A \cap B$ verbal im Kontext der Situation.

e) Beschreiben Sie die Menge $A \cup \bar{A}$ verbal im Kontext der Situation.

f) Welcher Anteil der Zuschauer, von denen man weiß, dass sie weiblich sind, hat das WM Finale angesehen?

g) Sind die Ereignisse A und B statistisch unabhängig (mit Begründung)?

2. Sie wissen, dass ein 40 prozentiger Anteil der Kinder das WM Finale angesehen hat. Wie groß ist die Wahrscheinlichkeit, dass von 20 unabhängig voneinander ausgewählten Kindern genau 13 das WM Finale angesehen haben?

3. Sie veranstalten Public Viewings während der Fußball WM und möchten Rückschlüsse auf den Getränkekonsum der Zuschauer ziehen. Dazu betrachten Sie folgende Zufallsvariable:

   X : „Verkaufte Getränkemenge in Liter während eines Public Viewings".

   Sie gehen davon aus, dass X normalverteilt ist und berechnen die Stichprobenvarianz von einer einfachen Stichprobe mit Stichprobenumfang $n = 5$ als $s_5^2 = 2.5$. Nun möchten Sie testen ob die Varianz kleiner als 2 ist und stellen dazu folgendes Hypothesenpaar auf:

$$H_0 : \sigma^2 \geq 2 \quad \text{vs.} \quad H_1 : \sigma^2 < 2.$$

a) Geben Sie die zugehörige Teststatistik und deren theoretische Verteilung an.

b) Berechnen Sie den Wert der Teststatistik.

c) Berechnen Sie den p-Wert. Falls Sie Teilaufgabe b) nicht lösen konnten, verwenden Sie den Wert $t_5 = 3$. Außerdem haben Sie die folgende Werte gegeben:

$$F_{\chi^2(5)}(5) = 0.5841,\ F_{\chi^2(4)}(5) = 0.7127, F_{\chi^2(4)}(4) = 0.5940,\ F_{\chi^2(5)}(4) = 0.4506,$$
$$F_{\chi^2(4)}(3) = 0.4422,\ F_{\chi^2(5)}(3) = 0.3000,\ F_{\chi^2(3)}(4) = 0.7385$$

d) Treffen Sie eine Testentscheidung mit einer Irrtumswahrscheinlichkeit von 10% (mit Begründung). Falls Sie Teilaufgabe c) nicht lösen konnten, verwenden Sie 0.5 als p-Wert.

4. Die Verteilung einer Zufallsvariable ist gegeben durch $X \sim \mathcal{N}(20, 4)$.

a) Berechnen Sie die Wahrscheinlichkeit des 3-fachen zentralen Schwankungsintervalls für X.

b) Geben Sie das 70% Quantil von X an.

Schmierpapier zu Aufgabe 1

Aufgabe 2 von 4

In einer Bank werden die einzelnen Kreditnehmer verschiedenen Ratingklassen zugeordnet, wobei AAA die beste und C die schlechteste Ratingkategorie beschreibt. Das Rating eines Kreditnehmers sei durch die Zufallsvariable X beschrieben:

Ratingklasse x	AAA	AA	A	BBB	BB	B	C
$P(X = x)$	0.08	0.12	0.28	0.18	0.11	0.08	θ

1. Welchen Wert muss θ annehmen, damit eine Wahrscheinlichkeitsverteilung vorliegt?

2. Welche Ausprägungen nehmen Median und Modus von X an?

3. Berechnen Sie die Gini Entropie für X. (Hinweis: Falls Sie Teilaufgabe 1 nicht lösen konnten, verwenden Sie $\theta = 0.1$)

Im Folgenden werden die Ratingklassen zweier Kreditnehmer X_1 und X_2 betrachtet, die beide jeweils der oben angegebenen Verteilung folgen.

4. Berechnen Sie die Wahrscheinlichkeit $P(X_1 = AAA|X_2 = AAA)$, dass Kreditnehmer X_1 Ratingklasse AAA besitzt, gegeben dass Kreditnehmer X_2 bereits in Kategorie AAA eingestuft wurde, wenn $P(X_1 = AAA \cap X_2 = AAA) = 0.05$ ist.

5. Wie groß ist die Wahrscheinlichkeit $P(X_1 = AAA|X_2 = AAA)$ unter der Voraussetzung, dass X_1 und X_2 stochastisch unabhängig sind?

Im Folgenden sind Realisationen der täglichen Rendite R_i einer Aktie von 10 aufeinander folgenden Tagen gegeben.

Tag i	1	2	3	4	5	6	7	8	9	10
Tägliche Rendite r_i	0.04	-0.04	0.04	0.02	-0.03	0.02	0.01	-0.01	-0.03	0.00

6. Geben Sie sowohl Merkmalstyp als auch maximales Skalenniveau des Merkmals "Tägliche Rendite" an.

7. Der Wert der Aktie zu Beginn des 10-Tages-Zeitraums beträgt 75 Euro, welchen Wert hat die Aktie zum Ende des 10-Tages-Zeitraums?

Nehmen Sie nun an, dass die Renditen R_i unabhängig und identisch normalverteilt sind mit Parametern μ_R und σ_R^2.

8. Berechnen Sie das Stichprobenmittel $\bar{R}_{10}$! Hinweis: $\sum_{i=1}^{10} r_i = 0.02$

9. Berechnen Sie die Stichprobenvarianz S_{10}^2! Hinweis: $\sum_{i=1}^{10} r_i^2 = 0.0076$

10. Nehmen Sie nun an, dass $\bar{r}_{10} = 0.001$ und $S_{10}^2 = 0.0006$ ist. Berechnen Sie das realisierte, symmetrische 95% Konfidenzintervall für den Mittelwert.

11. Gehen Sie nun davon aus, dass das realisierte 95% Konfidenzintervall für den Mittelwert durch $[-0.004, 0.007]$ gegeben ist. Begründen Sie kurz, ob Sie die zugehörige Hypothese H_0: $\mu_R = 0$ verwerfen können oder nicht.

12. Nehmen Sie nun an, dass die Momente des Stichprobenmittels $\bar{R}$ bekannt sind, wobei $\mu_{\bar{R}} = 0$ und $\sigma_{\bar{R}}^2 = 0.0009$. Bestimmen Sie die Grenzen eines zentralen Schwankungsintervalls in dem $\bar{R}$ mit 95% Wahrscheinlichkeit enthalten ist mit Hilfe der Chebyshevschen Ungleichung.

Schmierpapier zu Aufgabe 2

Aufgabe 3 von 4

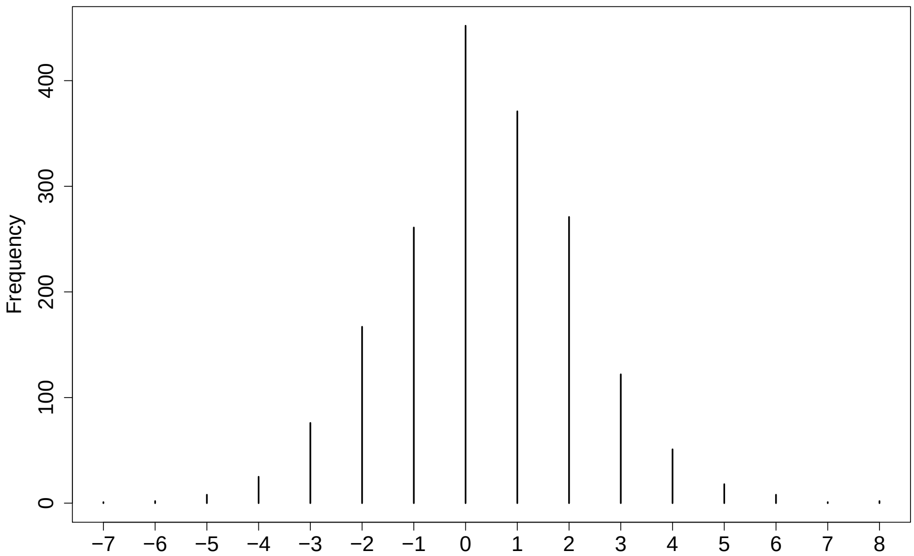
In Ihrem Workspace liegt ein Dataframe `data`, welcher folgende Informationen zu 1836 Fußballspielen enthält:

Spaltenname	**Info**
`Season`	Saison
`Home`	Heimmannschaft
`Away`	Auswärtsmannschaft
`Goal_h`	Tore Heimmannschaft
`Goal_a`	Tore 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 an, mit dem Sie die durchschnittliche Anzahl an Toren der Auswärtsmannschaft erhalten.

2. Geben Sie den `R`-Code für die Berechnung der durchschnittlichen Tore der Heimmannschaft an, bedingt darauf, dass eine Heimmannschaft mindestens 1 Tor schießt.

3. Welches der folgenden Spielergebnisse wird am häufigsten beobachtet: Sieg der Heimmannschaft, Unentschieden, Sieg der Auswärtsmannschaft?

4. Welches der folgenden beiden Spielergebnisse tritt häufiger auf: Sieg der Heimmannschaft mit einem Tor Vorsprung oder Sieg der Auswärtsmannschaft mit einem Tor Vorsprung. Eine Begründung ist erforderlich!

5. Geben Sie den `R`-Code an, mit dem Sie alle Daten aus der Saison 2017-2018 erhalten.

6. Geben Sie den `R`-Code an, mit dem Sie die Anzahl der Spiele erhalten, bei denen kein Tor gefallen ist.

7. Welchen Output liefert der Befehl `data$Home[seq(1,4,2)]`?

8. Mit welchem `R`-Code können Sie sich den Teildatensatz `teil1` abspeichern, der nur Fußballspiele enthält, welche `Moenchengladbach` als Heimmanschaft besitzen?

In Ihrem Workspace finden Sie nun folgende Funktionen:

```
f1=function(x)
{
tab = table(x)
max = max(tab)
element=which(tab==max)
names(element)
}

f2=function(x)
{
mean((x-mean(x))^2)
}
```

9. Welche statistische Kennzahl wird mithilfe von `f1` berechnet? Welche statistische Kennzahl wird mithilfe von `f2` berechnet?

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

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

10. Welchen Wert erhalten Sie für j?

11. Vervollständigen Sie den folgenden `R`-Code, um eine Funktion zu erhalten mit der man den Korrelationskoeffizient zwischen den Vektoren `x` und `y` erhält:

```
f3 = function(x,      ){
 1/(sd(y)*      (x))*      (x,y)
}
```

12. Mit welchem `R`-Code können Sie den Korrelationskoeffizient zwischen Tore Heimmannschaft und Tore Auswärtsmannschaft bestimmen? Nutzen Sie hierzu Funktion `f3`.

Schmierpapier zu Aufgabe 3

Aufgabe 4 von 4

Sie besitzen einen Kiosk.

1. Die Variable P: 'Anzahl Personen m, die in 5 Minuten Ihren Kiosk betreten' stamme aus einer Poissonverteilung, wobei die Annahmen an einen Poissonprozess erfüllt sind. Sie stoppen nun die Zeit in Minuten zwischen dem Ankommen zweier Personen (Z).

a) Welcher Verteilungsfamilie folgt die Variable Z (ohne Parameter)?

b) Beschreiben Sie allgemein die Methode der Momente.

c) Schätzen Sie den Parameter der Verteilung von Z mit Hilfe von $\overline{z}_m = 0.4$.

d) Berechnen Sie die Varianz der Verteilung von Z und verwenden Sie dafür $\lambda = 2.2$.

2. Jede Woche bestellen Sie von zwei unterschiedlichen Anbietern (Anbieter A und Anbieter B) Bonbons für Ihren Kiosk. Sie haben die Bonbons in den Geschmacksrichtungen süß und sauer bestellt. Sie wissen, dass die Wahrscheinlichkeit für ein saures Bonbon, wenn das Bonbon von Anbieter A kommt 30% und wenn das Bonbon von Anbieter B kommt 50% ist.

a) Welchen Merkmalstyp hat das Merkmal G: 'Geschmacksrichtung eines Bonbons'?

b) Wie ist das Merkmal G: 'Geschmacksrichtung eines Bonbons' skaliert?

c) Wenn Sie gleich viele Bonbons von Anbieter A und Anbieter B in einer Schüssel zusammenschütten, wie hoch ist dann die Wahrscheinlichkeit ein saures Bonbon daraus zu ziehen?

d) Wieviele Bonbons müssen Sie mindestens aus einer der Packungen ziehen, dass mit einer Irrtumswahrscheinlichkeit von 5% der Anteil der sauren Bonbons in Ihrer Stichprobe zur Grundgesamtheit um weniger als 10% vom wahren Anteil nach oben und unten abweicht? Nutzen Sie die Ungleichung von Chebyshev.

3. Nun haben Sie die Vermutung, dass ein Zusammenhang zwischen der Dauer in Minuten, die Ihre Kunden in Ihrem Kiosk verbringen (M) und dem Betrag in Euro, den Ihre Kunden bei Ihnen ausgeben (B) besteht. Sie haben eine einfache Stichprobe vom Umfang $n = 25$ gezogen und notieren sich die jeweiligen Realisationen von M und B.

Sie vermuten, dass M und B korreliert sind.

a) Was bedeutet Korrelation?

Testen Sie Ihre Vermutung bei einer Irrtumswahrscheinlichkeit von 10%, wenn Sie wissen, dass (M,B) bivariat normalverteilt ist mit unbekanntem Erwartungswertvektor (μ_M, μ_B) und unbekanntem Varianzvektor (σ_M^2, σ_B^2).
Sie stellen folgende Hypothesen auf:

$$H_0 : \rho = 0 \qquad H_1 : \rho \neq 0$$

Sie haben außerdem gegeben, dass der Wert des Stichprobenkorrelationskoeffizienten $R = 0.7$ ist. Gehen Sie davon aus, dass die Approximationsbedingungen erfüllt sind.

b) Berechnen Sie den Wert der Teststatistik.

c) Wie lautet der Ablehnungsbereich mit konkretem Wert?

d) Nehmen Sie nun an, dass der Hypothesentest einen p-Wert von 0.01 liefert. Treffen Sie eine Testentscheidung auf dem 10% Signifikanzniveau.

Nun wollen Sie das approximative Konfidenzintervall für den Korrelationskoeffizienten ρ bilden.

e) Berechnen Sie hierfür zunächst Fishers Z-Transformation.

Verwenden Sie ab jetzt: $Z(R) = 0.8$

f) Berechnen Sie nun die untere Grenze des approximativen realisierten 90%-Konfidenzintervalls für ρ.

Schmierpapier zu Aufgabe 4

Lösung Klausur SoSe 18 (10 ECTS)

Lösung 1

1a)	0.15	0.5 P
1b)	0.65	0.5 P
1c)	0.6	0.5 P
1d)	Menge der Personen, die weiblich sind und das WM Finale gesehen haben.	0.5 P
1e)	Menge aller Fernsehzuschauer.	0.5 P
1f)	$P(B\|A) = 0.625$	1 P
1g)	$P(A \cap B) = 0.25 \neq P(A)P(B) = 0.14 \Rightarrow$ Nicht statistisch unabhängig.	1 P
2.	0.0146	0.5 P
3a)	$(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$	1 P
3b)	5	0.5 P
3c)	$p = 0.7127$	1 P
3d)	$p > \alpha \Rightarrow$ nicht ablehnen.	1 P
4a)	$P(-3 < \frac{X-\mu}{\sigma} < 3) = 0.9974$	1 P
4b)	21.0488	0.5 P

Lösung 2

1. $\theta = 0.15$ 0.5 P

2. $\bar{x}_{\text{mod}} =$ "A", $\bar{x}_{\text{med}} =$ "BBB" 1 P

3. $0.08 \cdot 0.92 + 0.12 \cdot 0.88 + \ldots + 0.15 \cdot 0.85 = 0.8274$
   Mit Ersatzergebnis: $\sum_i p_i(1-p_i) = 0.7899$, $1 - \sum_i p_i^2 = 0.8399$ 1 P

4. $P(X_1 = AAA|X_2 = AAA) = \frac{P(X_1=AAA \cap X_2=AAA)}{P(X_2=AAA)} = \frac{0.05}{0.08} = 0.625$ 1 P

5. $P(X_1 = AAA|X_2 = AAA) = P(X_1 = AAA) = 0.08$ 0.5 P

6. Merkmalstyp: quantitativ, Skalenniveau: verhältnis-/absolut skaliert 1 P

7. $75[EUR] \cdot \prod_{i=1}^{10}(1+r_i) = 76.23[EUR]$ 0.5 P

8. $\bar{R}_{10} = \frac{\sum_{i=1}^{10} r_i}{10} = 0.002$ 0.5 P

9. $S_{10}^2 = \frac{1}{9}\left(\sum_{i=1}^{10} r_i^2 - 10\bar{r}^2\right) = \frac{1}{9}\left(\sum_{i=1}^{10} r_i^2 - \frac{1}{10}\left(\sum_{i=1}^{10} r_i\right)^2\right)$
   $= 0.0008$ 1 P

10. $\left[\bar{r}_{10} - t_{1-\frac{\alpha}{2};9}\frac{S_{10}}{\sqrt{10}}; \bar{r}_{10} + t_{1-\frac{\alpha}{2};9}\frac{S_{10}}{\sqrt{10}}\right] = [-0.0165; 0.0185]$
    wobei $t_{1-\frac{\alpha}{2};9} = 2.262$ 1.5 P

11. H_0 beibehalten, da 0 innerhalb der Intervallgrenzen $[-0.004; 0.007]$. 0.5 P

12. $P(\mu_R - k \cdot \sigma_R < R < \mu_R + k \cdot \sigma_R) \geq 1 - \frac{1}{k^2} = 0.95 \Rightarrow k = \sqrt{20}$
    $\Rightarrow P(-0.1342 < R < 0.1342) = 0.95 \Rightarrow g_u = -0.1342, g_o = 0.1342$ 1 P

Lösung 3

1. `mean(data$Goal_a)` 0.5 P
2. `mean(data$Goal_h[data$Goal_h>=1])` 1 P
3. Heimmannschaft 0.5 P
4. Sieg der Heimmanschaft mit einem Tor Vorsprung, da absolute Häufigkeit größer. 0.5 P
5. `data[data$Season=="2017-2018",]` 1 P
6. `sum(data$Goal_h==0&data$Goal_a==0)` 1 P
7. Borussia Dortmund, FC Augsburg 1 P
8. `teil1 = data[data$Home == "Moenchengladbach",]` 1 P
9. Modus, Varianz 1 P
10. 7 0.5 P
11. y,sd,cov 1.5 P
12. `f3(data$Goal_h,data$Goal_a)` 0.5 P

Lösung 4

1. a) Die Zeit zwischen dem Eintreten zweier Poisson-Ereignisse ist exponentialverteilt.
   0.5 P
   b) Methode der Momente: Potenzmomente der Stichprobe werden mit entsprechenden Potenzmomenten der Grundgesamtheit gleichgesetzt. 0.5 P
   c) $\hat{\lambda} = 1/\overline{Z}_m = 2.5$ 0.5 P
   d) Varianz: $1/\lambda^2 = 0.2066$. 0.5 P
2. a) Merkmalstyp: qualitativ 0.5 P
   b) Skalierung: Nominal 0.5 P
   c) 0.4 0.5 P
   d) $1/(4 * 0.05 * 0.1^2) = 500$ 1 P
3. a) Korrelation: (linearer) Zusammenhang zweier Zufallsvariablen 0.5 P
   b) Teststatistik: 4.7009 1 P
   c) Ablehnungsbereich: $|t_n| > t_{1-\alpha/2;n-2}$, wobei $t(0.95; 23) = 1.714$ 1 P
   d) Testentscheidung: Da $p = 0.01 < \alpha = 0.1$ ($|4.7009| > 1.714$) 0.5 P
   Die Nullhypothese kann auf dem 10% Signifikanzniveau abgelehnt werden.
   0.5 P
   e) Fishers Z-Transformation: Z(R)=0.8673 0.5 P
   f) $[\frac{e^{2(Z(R)-\lambda_{1-\alpha/2}\frac{1}{\sqrt{n-3}})}-1}{e^{2(Z(R)-\lambda_{1-\alpha/2}\frac{1}{\sqrt{n-3}})}+1}; \quad \cdot \quad]$ 0.5 P

   $\lambda_{1-\alpha/2} = 1.6448$ 0.5 P
   $[0.4213; \quad \cdot \quad]$ 0.5 P