Klausur Statistik (7,5 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 05.08.2015
Matrikelnummer			14:00 - 16:00 Uhr
Studienrichtung		**Sitzplatznummer**	
Semesterzahl		**Raum**	
Email (optional)			

Hinweise: 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 und vom Lehrstuhl autorisierte Fehlerkorrekturen
- 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

Bei einem Golfturnier in der Golfsportanlage SANDBUNKER wurden die 32 teilnehmenden Spieler auf Grundlage ihres Handicaps in zwei Gruppen (Merkmal G) eingeteilt: In Gruppe G_1 befinden sich die Spieler mit niedrigem und in Gruppe G_2 die Spieler mit höherem Handicap. Beim ersten Loch benötigten die Spieler eine gewisse Anzahl an Schlägen (Merkmal S). In folgender Tabelle sind die relativen Häufigkeiten f_{ij} und die relativen Randhäufigkeiten $f_{i\cdot}$ bzw. $f_{\cdot j}$ aufgeführt.

$i\backslash j$		1	2	
	$S\backslash G$	G_1	G_2	$f_{i\cdot}$
1	S_1: 2 Schläge	0,0625	0,03125	0,09375
2	S_2: 3 Schläge	0,125	0,03125	0,15625
3	S_3: 4 Schläge	0,15625	0,125	
4	S_4: 5 Schläge	0,21875	0,25	0,46875
	$f_{\cdot j}$	0,5625	0,4375	1

1. Ergänzen Sie den fehlenden Wert $f_{3\cdot}$ in der Tabelle.

2. Geben Sie zu Merkmal G Modus und Modalhäufigkeit an und interpretieren Sie letztere.

3. Berechnen Sie die Gini-Entropie für Merkmal G. Interpretieren Sie das Ergebnis; geben Sie dazu auch die Obergrenze an.

4. Berechnen Sie die absolute Häufigkeit n_{21} und interpretieren Sie diese.

5. Wie viele Teilnehmer benötigten 5 Schläge?

6. Sind die beiden Merkmale S und G statistisch unabhängig? Begründen Sie Ihre Antwort.

7. Wie groß ist der Anteil der Spieler, die das erste Loch mit 2 Schlägen beenden konnten unter denen, die in die Gruppe 2 eingeteilt wurden?

8. Berechnen Sie die durchschnittliche Schlagzahl unter den Spielern in Gruppe 1.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Eine Fluggesellschaft bietet in einer Boeing B777-300 Linienflüge zwischen München und der Kapverdischen Insel Sal an. Pro Flug können 300 Passagiere befördert werden.

1. Zur Untersuchung der Effizienz des Flugplans wurde die Auslastung (in Prozent) des Flugzeugs im letzten Monat (16 Flüge) untersucht.

Nr. des Fluges	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Auslastung in %	78	85	93	97	84	88	72	90	60	89	91	75	71	82	98	99

Welcher Mittelwert ist zur Bestimmung der mittleren prozentualen Auslastung geeignet?

2. Erfahrungsgemäß wird in der Business-Class, die aus 20 Sitzen besteht, ein gebuchter Platz nur mit einer Wahrscheinlichkeit von 90% tatsächlich in Anspruch genommen. Sei

S : „Anzahl der tatsächlich belegten Plätze".

Dann gilt: $S \sim Bin(20, 0.9)$.
Bestimmen Sie die Wahrscheinlichkeit, dass

a) genau 4 der 20 Sitzplätze nicht belegt werden,

b) mindestens 12 der 20 Sitzplätze belegt werden,

c) weniger als 18 der 20 Sitzplätze belegt werden.

Leider ist der Fluggesellschaft die Rücktrittswahrscheinlichkeit p für gebuchte Sitzplätze in der Economy-Class nicht bekannt. Um dennoch eine Aussage über p treffen zu können, betrachtet ihre Controlling-Abteilung $N = 2500$ stochastisch unabhängige Buchungen in der Economy-Class. Sei

$$X_i = \begin{cases} 1 & i\text{-te Buchung wird } \textbf{nicht} \text{ angetreten} \\ 0 & i\text{-te Buchung wird angetreten} \end{cases}$$

für $i = 1, ..., 2500$. Sei

$R:$ „Anzahl der nicht angetretenen Buchungen"

mit $R = \sum_{i=1}^{N} X_i$ binomialverteilt mit Parametern $n = 2500$ und unbekannter Rücktrittswahrscheinlichkeit p.

3. Wie ist X_i $(i = 1, ..., 2500)$ verteilt?

Es sei nun bekannt, dass insgesamt $r = 243$ der gebuchten Flüge nicht angetreten wurden.

4. a) Bestimmen Sie das theoretische und das realisierte approximative 92.8%- Konfidenzintervall für die Rücktrittswahrscheinlichkeit p.

   b) Interpretieren Sie das realisierte Konfidenzintervall aus Teilaufgabe 4.a).

5. Die Fluggesellschaft vermutet, dass ein gebuchter Platz in der Economy-Class mit einer geringeren Wahrscheinlichkeit nicht besetzt wird als in der Business-Class. Dazu führt sie auf dem 92.8%-Niveau einen approximativen Hypothesentest mit dem Hypothesenpaar

$$H_0: \; p \geq 0.1 = p_0 \qquad\qquad H_1: \; p < 0.1$$

durch. Die zugehörige Prüfgröße lautet

$$T_N = \sqrt{N}\frac{\overline{X}_N - p_0}{\sqrt{p_0(1-p_0)}} \overset{asy}{\sim} \mathcal{N}(0,1)$$

a) Bestimmen Sie die realisierte Prüfgröße. Runden Sie in dieser Teilaufgabe auf zwei Nachkomma-Stellen genau.

b) Bestimmen Sie die kritische Schranke.

c) Berechnen Sie den p-Wert. Sollten Sie bei Teilaufgabe 5.a) kein Ergebnis erhalten haben, verwenden Sie $t_N = -0.48$.

d) Treffen Sie eine Testentscheidung und begründen Sie diese.

Da die Fluggesellschaft Teilhaberin an einem Hotel auf einer der Nachbarinseln ist, bietet sie einen Transfer-Flug in einer kleinen Charter-Maschine (23 Sitzplätze) von der Insel Sal aus an. Um die Auslastung dieser Maschine möglichst hoch zu halten, ist die Fluggesellschaft dazu übergegangen, diese Flüge zu überbuchen und insgesamt 25 Buchungen zuzulassen.

6. Mit welcher Wahrscheinlichkeit muss die Fluggesellschaft mindestens einen Passagier zurückweisen, wenn bekannt ist, dass ein gebuchter Sitzplatz mit einer Wahrscheinlichkeit von 10% nicht belegt wird?
   (Hinweis: Sei $R_{Charter}$: „Anzahl der nicht belegten Sitzplätze" binomialverteilt mit Parameter $n = 25$ und $p = 0.1$.)

Schmierpapier zu Aufgabe 2

Aufgabe 3

Teil I:

1. Eine Investition hat zum Zeitpunkt t_0 einen Wert von 100 EUR. In den darauf folgenden zehn Jahren beträgt die durchschnittliche Wachstumsrate 3.2 Prozent. Welchen Wert hat die Investition nach Ablauf der zehn Jahre?

2. Sie beobachten folgende Preisentwicklung einer Aktie A über fünf Jahre. Angegeben ist jeweils der Preis in EUR zum Jahresende.

Jahr i	1	2	3	4	5
Preis Aktie A in EUR	100	90	130	150	140

Ein Anleger kauft am Ende von Jahr 1 insgesamt zehn Aktien. Wie hoch ist seine durchschnittliche Rendite pro Jahr am Ende von Jahr 5?

3. In drei aufeinanderfolgenden Jahren kauft ein Anleger Anteile einer Aktie. Er investiert 800 EUR im ersten Jahr, 600 EUR im zweiten Jahr und 200 EUR im dritten Jahr. Der Preis pro Anteil beträgt 25 EUR im ersten Jahr und 30 EUR im zweiten Jahr. Ferner ist über die drei Jahre ein durchschnittlicher Preis pro Anteil von 28.07 EUR bezahlt worden. Wie hoch ist der Preis pro Anteil im dritten Jahr?

Teil II:

4. An den Finanzmärkten treten immer wieder Extremereignisse auf, also besonders hohe oder niedrige Renditen. Ihnen sei folgende Zufallsvariable gegeben:

   X: „Anzahl der eingetretenen Extremereignisse pro Jahr“

   Ein Börsenjahr bestehe aus 250 Tagen. Es ist bekannt, dass im Mittel zwei Extremereignisse pro Jahr auftreten. Nehmen Sie an, dass X Poisson-verteilt ist.

   a) Nennen Sie die drei Annahmen des Poisson-Prozesses (nur verbal).

   b) Welcher Verteilung folgt die Wartezeit W zwischen zwei Poisson-Ereignissen?

   c) Berechnen Sie die Fläche unter der Dichtefunktion der Exponentialverteilung mit Parameter λ allgemein bis zu einem Punkt $x \geq 0$ (Rechenweg erforderlich).

   d) Wie hoch ist die Wahrscheinlichkeit, dass die Wartezeit W zwischen zwei Extremereignissen mehr als zwei Jahre (= 500 Tage) beträgt?

Teil III:

5. Ihnen sei die Zufallsvariable Y: „Rendite des Anlageprodukts Alpha in Prozent“ gegeben. Die Zufallsvariable Y folge einer t-Verteilung mit 5 Freiheitsgraden.

   a) Geben Sie den Erwartungswert von Y an.

   b) Bestimmen Sie die Varianz von Y.

   c) Welche Rendite wird mit einer Wahrscheinlichkeit von 99.9% nicht unterschritten?

   d) Ihnen sei weiterhin die Zufallsvariable Z: „Rendite des Anlageprodukts Beta in Prozent“ gegeben. Die Zufallsvariable Z folge einer Normalverteilung mit $\mu = 0$ und $\sigma = 1.29$. Welche Rendite wird nun mit einer Wahrscheinlichkeit von 99.9% nicht unterschritten?

Schmierpapier zu Aufgabe 3

Aufgabe 4

Sie sind Lehrer an einer Wirtschaftsschule und haben eine neue Methode entwickelt, Schülern das Fach Wirtschaft zu lehren. Sie wissen, dass bei Wirtschaftsschulaufgaben im Mittel 14 Punkte mit einer Standardabweichung von 7 Punkten erreicht werden. Die Ergebnisse einer Schulaufgabe die von 35 Schülern geschrieben wurde, die Sie unterrichtet haben, sehen Sie in der folgenden Tabelle:

Punkte	**Anzahl Schüler** n_i
0 bis unter 5	3
5 bis unter 10	3
10 bis unter 15	7
15 bis unter 20	12
20 bis unter 25	6
25 bis unter 30	4

1. Berechnen Sie den Mittelwert der erreichten Punkte in dieser Schulaufgabe.

Hinweis: Falls Sie diese Aufgabe nicht lösen konnten, nehmen Sie im Weiteren einen Mittelwert von 16.5 Punkten an.

Sie haben die Vermutung, dass Schüler, die mit Ihrer Methode unterrichtet wurden, bessere Ergebnisse in Schulaufgaben erzielen.

2. Um diese Vermutung zu überprüfen, testen Sie zuerst bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$, ob Ihre Daten normalverteilt sind. Dazu haben Sie sich folgende Tabelle als Arbeitshilfe (auf 4 Nachkomma-Stellen gerundet) erstellt:

Punkte	p_i^0	np_i^0
$-\infty$ bis unter 5	0.0980	3.4300
5 bis unter 10	0.1831	6.4085
10 bis unter 15	0.2729	
15 bis unter 20		8.6631
20 bis unter 25	0.1365	4.7985
25 bis $+\infty$	0.0620	2.1700

a) Vervollständigen Sie die Angaben in der Arbeitstabelle.

Sie fassen die erste und zweite Kategorie sowie die vorletzte und letzte zu je einer Kategorie zusammen.

b) Nennen Sie den Grund für dieses Vorgehen.

c) Bestimmen Sie die kritische Schranke für Ihren Test auf Vorliegen einer Normalverteilung bei Ihren Daten.

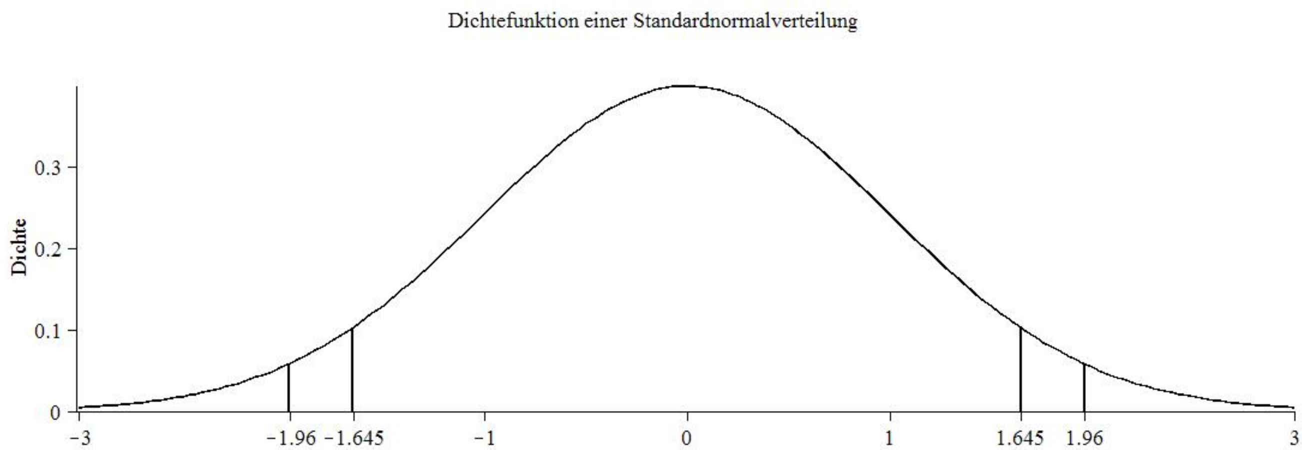
Arbeiten Sie im Weiteren mit der Normalverteilungsannahme für Ihre Daten. Die Standardabweichung der Grundgesamtheit gilt weiterhin als bekannt mit $\sigma = 7$.

3. Nun wollen Sie die Vermutung überprüfen, dass Ihre neue Lehrmethode signifikant höhere durchschnittliche Punktzahlen in Wirtschaftsschulaufgaben liefert. Somit ist:

$$H_0 : \mu \leqslant 14 \quad \text{gegen} \quad H_1 : \mu > 14$$

a) Erklären Sie die Bedeutung des Begriffs "signifikant".

b) Der folgende Plot zeigt die Dichtefunktion einer Standardnormalverteilung mit $\mu = 0$ und $\sigma^2 = 1$. Markieren Sie den Ablehnungsbereich mit der Irrtumswahrscheinlichkeit von $\alpha = 5\%$ für das Hypothesenpaar in diesem Plot eindeutig.



c) Berechnen Sie die realisierte Prüfgröße.

Falls Sie die letzte Aufgabe nicht lösen konnten, verwenden Sie als Wert der realisierten Prüfgröße: $t_n = 2.00$.

d) Berechnen Sie den p-Wert und treffen Sie eine Testentscheidung.

Nehmen Sie im Folgenden an, dass ein signifikanter Punktunterschied zwischen den Ergebnissen der Schüler, die Sie mit Ihrer neuen Lehrmethode unterrichteten und den allgemeinen Ergebnissen von Wirtschaftsschulaufgaben vorliegt.

4. Nun interessiert es Sie, ob dies ein großer oder vernachlässigbar kleiner Unterschied ist.

a) Berechnen Sie dazu den Schätzer der Effektgröße Cohen's δ:

Nehmen Sie an, Sie erhalten als Effektgröße $\delta = 0.3$.

b) Interpretieren Sie diesen Effekt anhand von Cohen's Interpretations-Grenzen Vorschlag.

5. Sie unterhalten sich mit dem Direktor der Schule über Ihre Ergebnisse. Dieser fragt Sie nach der Wahrscheinlichkeit des Fehlers 2. Art Ihres Tests.

a) Erklären Sie was der Fehler 2. Art aussagt.

Sie berechnen die Wahrscheinlichkeit des Fehlers 2. Art für den in 3. durchgeführten Hypothesentest und erhalten $\beta = 0.36$.

b) Berechnen Sie die Güte des Tests.

c) Nennen Sie eine Möglichkeit, wie die Güte des Tests erhöht werden kann.

d) Berechnen Sie den Stichprobenumfang, den Sie brauchen um eine Testentscheidung für den in 3. durchgeführten Hypothesentest mit einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ und einer Güte von 80% für $\mu_1 = 16.5$ treffen zu können.

Schmierpapier zu Aufgabe 4

Lösung Klausur SS15 (7.5ECTS)

Aufgabe 1

1. 0.28125. 0.5P

2. Modus G_1 Modalhäufigkeit 0.5625
56,25% der Teilnehmer wurden der Gruppe der Spieler zugeordnet, die ein niedrigeres Handicap haben. 1.5P

3. $H_2 = 1 - 0.5625^2 - 0.4375^2 = 0.4922$
Obergrenze: 1/2 hohe Streuung 2P

4. $n_{2,1} = 32 \cdot 0.125 = 4$
4 Personen wurden der Gruppe 1 zugeordnet und benötigten drei Schläge. 1P

5. 15 Personen brauchten 5 Schläge. 0.5P

6. Nein, z.B. $0,09375 \cdot 0,5625 = f_{1,\cdot} \cdot f_{\cdot,1} \neq f_{1,1} = 0{,}0625$ 1.5P

7. $f(s_1|g_2) = \frac{f(s_1 \cap g_2)}{f(g_2)} = \frac{0{,}03125}{0{,}4375} = 0{,}0714$ 1P

8. Mit bedingten Häufigkeiten $f(S = S_i|G = G_1) = \frac{f_{i,1}}{f_{\cdot,1}}$:

S_i	S_1	S_2	S_3	S_4
$f(S = S_i \vert G = G_1)$	0.1111111	0.2222222	0.2777778	0.3888889

$$\begin{aligned}\bar{s}_{|(G=G_1)} &= 0.1111111 * 2 + 0.2222222 * 3 + 0.2777778 * 4 + 0.3888889 * 5 \\ &= 3,9444\end{aligned}$$

2P

Aufgabe 2

1. arithmetisches Mittel 0.5 P

2. a) $P(S=16) = P(20-S=4) = f_{Bin(20,0.1)}(4) = 0.0898$ 1 P

   b) $P(S \geq 12) = P(20-S \leq 8) = F_{Bin(20,0.1)}(8) = 0.9999$ 1 P

   c) $P(S < 18) = P(20-S \geq 3) = 1 - P(20-S \leq 2) = 1 - F_{Bin(20,0.1)}(2) = 1 - 0.6769 = 0.3231$ 1P

3. Bernoulli (mit Parameter p) 0.5 P

4. a) <u>theoretisches Konfidenzintervall:</u>

$$\left[\frac{R}{N} - \lambda_{0.964}\frac{\sqrt{\frac{R}{N}\left(1-\frac{R}{N}\right)}}{\sqrt{N}}\ ;\ \frac{R}{N} + \lambda_{0.964}\frac{\sqrt{\frac{R}{N}\left(1-\frac{R}{N}\right)}}{\sqrt{N}}\right]$$

0.5 P

<u>realisiertes Konfidenzintervall:</u>

$$\left[\frac{243}{2500} - 1.7991\frac{\sqrt{\frac{243}{2500}\left(1-\frac{243}{2500}\right)}}{\sqrt{2500}}\ ;\ \frac{243}{2500} + 1.7991\frac{\sqrt{\frac{243}{2500}\left(1-\frac{243}{2500}\right)}}{\sqrt{2500}}\right] = [0.0865\ ;\ 0.1079]$$

1 P

   b) Mit einer Vertrauenswürdigkeit von 92.8% liegt die Rücktrittswahrscheinlichkeit im obigen realisierten Konfidenzintervall. 0.5 P

   oder

   In 92.8% der Fälle liegt die tatsächliche Rücktrittswahrscheinlichkeit p im realisierten Konfidenzintervall.

5. a) $t_{2500} = \sqrt{2500}\frac{\frac{243}{2500} - 0.1}{\sqrt{0.1 \cdot 0.9}} = -0.4667 \approx -0.47$ 0.5 P

   b) $-\lambda_{0.928} = -1.4611$ 1 P

   c) $p - Wert = P(T_N \leq -0.47) \overset{asy}{\approx} \Phi(-0.47) = 1 - 0.6808 = 0.3192$ 1 P
   ($p - Wert = P(T_N \leq -0.48) \overset{asy}{\approx} \Phi(-0.48) = 1 - 0.6844 = 0.3156$)

   d) H_0 nicht ablehnen, da $t_N = -0.4667 \not< -1.4611 = -\lambda_{0.928}$ 0.5 P

   oder

   H_0 nicht ablehnen, da $p = 0.3192 \not< 0.072 = \alpha$ 0.5 P

   oder

   H_0 nicht ablehnen, da $p_0 = 0.1$ im realsierten Konfidenzintervall liegt 0.5 P

6. $P(R \leq 1) = F_{Bin(25,0.1)}(1) = 0.2712$ 1 P

Aufgabe 3

1. 137.02 — 0.5 P+0.5 P

2. 0.0878 — 0.5 P

3. 40.00 — 0.5 P+0.5 P

4. a) P1 Wahrscheinlichkeit, in einem Intervall der Länge h das Ereignis genau einmal zu beobachten ist proportional zur Länge des Intervalls, aber unabhängig von dessen Lage. — 0.5 P

   P2 Wahrscheinlichkeit, in einem sehr kleinen Intervall der Länge h mehr als ein Ereignis zu beobachten ist vernachlässigbar klein. — 0.5 P

   P3 Die Anzahl der Ereignisse in zwei Intervallen, welche sich nicht überlappen, sind voneinander unabhängig. — 0.5 P

   b) Exponentialverteilt — 0.5 P

   c) $F_X(x; \lambda) = 1 - e^{-\lambda x}$ — 0.5 P+0.5 P

   d) 0.0183 — 0.5 P+0.5 P

5. a) 0 — 0.5 P

   b) 5/3 — 0.5 P+0.5 P

   c) -5.893 — 0.5 P+0.5 P

   d) -3.986 — 0.5 P+0.5 P

Aufgabe 4

1. Mittelwert = 16.3571 — 0.5 P

2. a) $p_4^0 = 0.2475$ — 0.5 P
   $np_3^0 = 9.5515$ — 0.5 P

   b) Bedingung für gute Approximation: $np_i^0 \geqslant 5$ für k $\leqslant$ 8 — 0.5 P

   c) $\chi^2(k-1=3) = 7.81$ — 0.5 P

3. a) Signifikant = "nicht-zufällig" — 0.5 P

   b) nur Fläche auf der rechten Seite markiert — 0.5 P
   Richtige Schranke $\lambda_{0.95} = 1.6448(1.645)$ gewählt — 0.5 P

   c) $T_n = \sqrt{n}\frac{\overline{x}_n-\mu_0}{\sigma} = \sqrt{35}\frac{(16.3571-14)}{7} = 1.9921$
   $(\sqrt{35}\frac{(16.5-14)}{7} = 2.1129)$ — 0.5 P+0.5 P

   d) p-Wert = $P(T_n > 2.1129(1.9921)(2.00)) = 0.0174(0.0233)(0.0228)$ — 0.5 P
   $\Rightarrow$ Testentscheidung: Lehne Nullhypothese ab. — 0.5 P

4. a) Cohen's $\delta = \frac{16.3571-14}{7} = 0.3367$
   $(\frac{16.5-14}{7} = 0.3571)$ — 0.5 P

   b) Cohen's δ von 0.3 (0.3367) (0.3571) $\Longrightarrow$ kleiner Effekt — 0.5 P

5. a) Fehler 2. Art = Fehler die Nullhypothese nicht abzulehnen, — 0.5 P
   obwohl sie falsch ist — 0.5 P

   b) $1 - 0.36 = 0.64$ — 0.5 P

   c) - Erhöhung von α
   - Erhöhung des Stichprobenumfangs n
   (- Erhöhung der Effektgröße
   (Differenz zwischen μ_1 und μ_0 erhöhen, bzw Varianz verringern)) — 0.5 P

   d) $n = ((\lambda_{1-\alpha} - \lambda_\beta)\frac{\sigma}{\mu_1-\mu_0}))^2$
   $= ((1.6448 + 0.8416)\frac{7}{16.5-14})^2 = 48.4683(= 49)$ — 0.5 P + 0.5 P

Klausur Statistik (10 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 05.08.2015 14:00 - 16:00 Uhr
Matrikelnummer			
Studienrichtung		**Sitzplatznummer**	
Semesterzahl		**Raum**	
Email (optional)			

Hinweise: 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 und vom Lehrstuhl autorisierte Fehlerkorrekturen
- 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

Bei einem Golfturnier in der Golfsportanlage SANDBUNKER wurden die 32 teilnehmenden Spieler auf Grundlage ihres Handicaps in zwei Gruppen (Merkmal G) eingeteilt: In Gruppe G_1 befinden sich die Spieler mit niedrigem und in Gruppe G_2 die Spieler mit höherem Handicap. Beim ersten Loch benötigten die Spieler eine gewisse Anzahl an Schlägen (Merkmal S). In folgender Tabelle sind die relativen Häufigkeiten f_{ij} und die relativen Randhäufigkeiten $f_{i\cdot}$ bzw. $f_{\cdot j}$ aufgeführt.

$i\backslash j$		1	2	
	$S\backslash G$	G_1	G_2	$f_{i\cdot}$
1	S_1: 2 Schläge	0,0625	0,03125	0,09375
2	S_2: 3 Schläge	0,125	0,03125	0,15625
3	S_3: 4 Schläge	0,15625	0,125	
4	S_4: 5 Schläge	0,21875	0,25	0,46875
	$f_{\cdot j}$	0,5625	0,4375	1

1. Ergänzen Sie den fehlenden Wert $f_{3\cdot}$ in der Tabelle.

2. Geben Sie zu Merkmal G Modus und Modalhäufigkeit an und interpretieren Sie letztere.

3. Berechnen Sie die Gini-Entropie für Merkmal G. Interpretieren Sie das Ergebnis; geben Sie dazu auch die Obergrenze an.

4. Berechnen Sie die absolute Häufigkeit n_{21} und interpretieren Sie diese.

5. Wie viele Teilnehmer benötigten 5 Schläge?

6. Sind die beiden Merkmale S und G statistisch unabhängig? Begründen Sie Ihre Antwort.

7. Wie groß ist der Anteil der Spieler, die das erste Loch mit 2 Schlägen beenden konnten unter denen, die in die Gruppe 2 eingeteilt wurden?

8. Berechnen Sie die durchschnittliche Schlagzahl unter den Spielern in Gruppe 1.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Ein Pharmaunternehmen beauftragt Sie einen Datensatz auszuwerten. Dabei sollen Sie Unterschiede in der Güte zweier Schwangerschaftstests ('Preggyfit' und 'Babyplan') herausarbeiten. Bei beiden Tests wird der hCG-Wert bestimmt, jedoch auf unterschiedliche Art und Weise. Insgesamt wurden beide Tests mit 187 teilnehmenden Frauen durchgeführt und dabei folgende Daten erhoben:

Merkmal		**Spaltenname**
hCG-Wert:	Konzentrationswert des Hormons hCG	`hCG.Wert`
Befund Preggyfit:	Befund von 'Preggyfit'	`Befund.Preggy`
Befund Babyplan:	Befund von 'Babyplan'	`Befund.Baby`
Schwanger:	War eine Frau tatsächlich schwanger?	`Wahrer.Zustand`

Dieser Datensatz steht Ihnen in `R` als Data Frame `schwangerschaft` zur Verfügung. Sie finden zunächst am Arbeitsplatz folgenden `R`-Output vor:

```
> str(schwangerschaft)

'data.frame':        187 obs. of  4 variables:
 $ hCG.Wert      : num  94.4 105.1 61.7 54.9 59.4 ...
 $ Befund.Preggy : chr  "positiv" "negativ" "positiv" "negativ" ...
 $ Befund.Baby   : chr  "positiv" "positiv" "positiv" "positiv" ...
 $ Wahrer.Zustand: logi  TRUE TRUE FALSE FALSE FALSE FALSE ...

> attach(schwangerschaft)
```

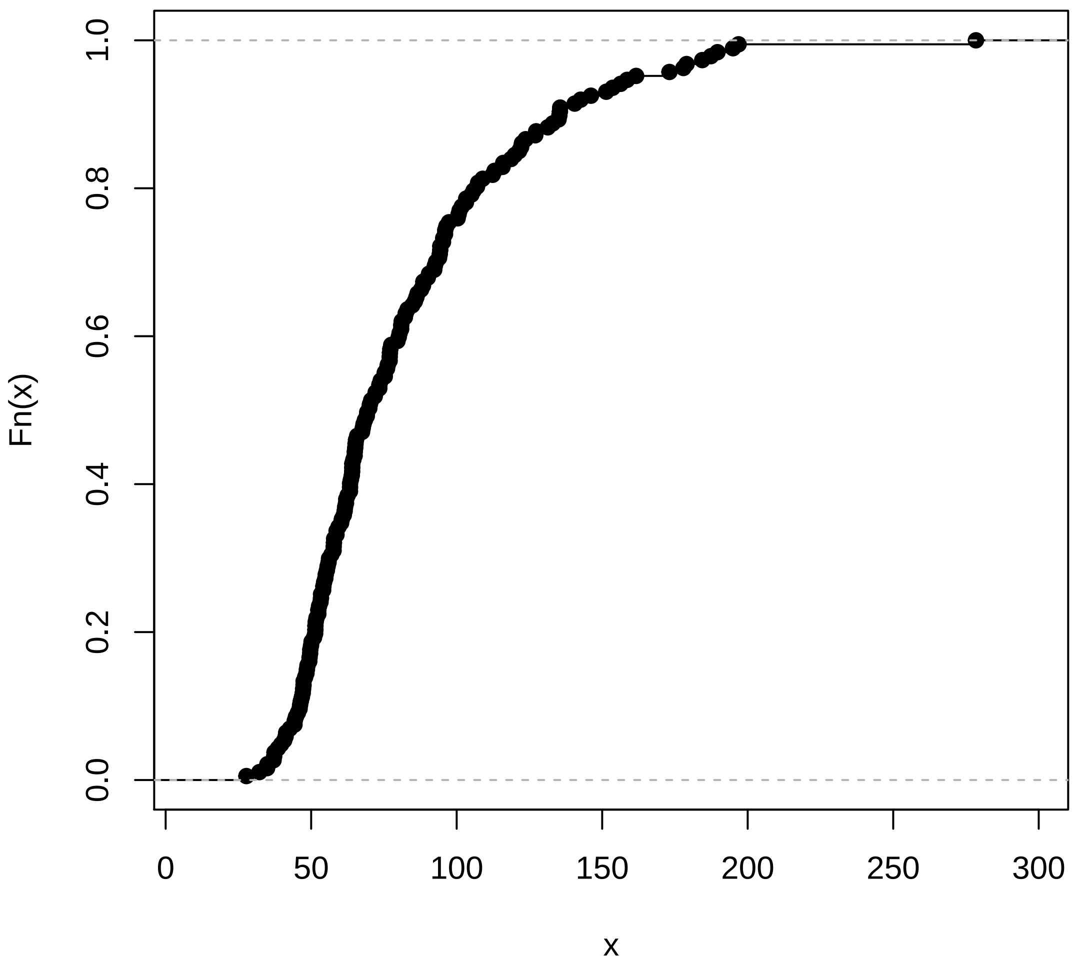
1. Wäre für das Merkmal *hCG-Wert* der Datentyp `character` geeignet gewesen? (**Kurze** Begründung)

2. Geben Sie den `R`-Befehl an, mit dem Sie sich das empirische 0%-Quantil des Merkmals *hCG-Wert* ausgeben lassen können.

3. Geben Sie den R-Befehl an, mit dem Sie hCG-Werte von Frauen angezeigt bekommen, die von Preggyfit oder Babyplan einen positiven Befund erhalten haben.

4. Geben Sie den R-Befehl an, um sich die empirische Verteilungsfunktion des Merkmals *hCG-Wert* graphisch anzeigen zu lassen.

Die folgende Grafik zeigt die empirische Verteilungsfunktion des Merkmals *hCG-Wert*.



5. Vervollständigen Sie die folgenden Aussagen über das Merkmal *hCG-Wert*:

   a) Das Maximum liegt ________________ dem Wert 250.

   b) Das Minimum ist ________________ 0.

Betrachten Sie nun zusätzlich folgende R-Befehle mit dem zugehörigen R-Output

```
> grenzen<-c(0,50,75,100,300)
> f1<-sum(hCG.Wert<grenzen[2])
> f2<-length(which(hCG.Wert>=grenzen[2] & hCG.Wert<grenzen[3]))
> f3<-length(which(hCG.Wert>=grenzen[3] & hCG.Wert<grenzen[4]))
> f4<-sum(hCG.Wert>=grenzen[4])
> c(f1,f2,f3,f4)/length(hCG.Wert)

[1] 0.1818182 0.3582888 0.2139037 0.2459893

> addmargins(table(Befund.Preggy,Befund.Baby))

             Befund.Baby
Befund.Preggy negativ positiv Sum
      negativ      44      58 102
      positiv      40      45  85
      Sum          84     103 187

> addmargins(table(Befund.Preggy,Wahrer.Zustand))

             Wahrer.Zustand
Befund.Preggy FALSE TRUE Sum
      negativ    68   34 102
      positiv    45   40  85
      Sum       113   74 187

> addmargins(table(Wahrer.Zustand,Befund.Baby))

         Befund.Baby
Wahrer.Zustand negativ positiv Sum
         FALSE      55      58 113
         TRUE       29      45  74
         Sum        84     103 187
```

6. Wie viel Prozent der untersuchten Frauen waren tatsächlich schwanger, wenn der Schwangerschaftstest Preggyfit einen positiven Befund lieferte?

7. Wie viele der untersuchten Frauen hatten einen *hCG-Wert* unter 50?

8. Berechnen Sie ausgehend von den verfügbaren Informationen den durchschnittlichen *hCG-Wert* einer untersuchten Frau. *Verwenden Sie dazu die Klassenmitte als Repräsentanten.*

9. Geben Sie den `R` Code an, um die Funktion

$$g(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2}$$

in `R` zu implementieren:

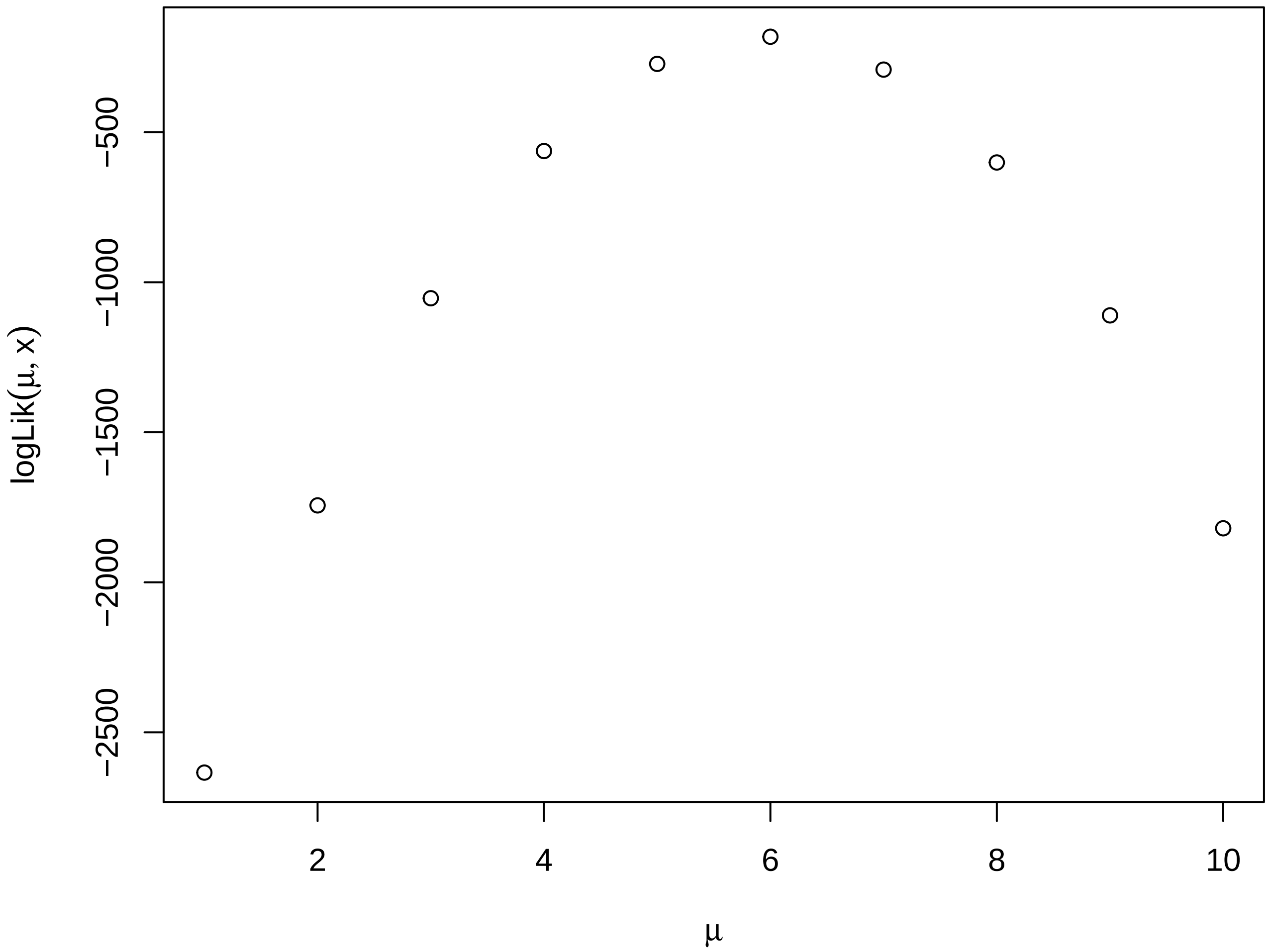
```
g=function(   ,mu){
      sqrt(2*pi)^(-1)*
}
```

Sie möchten nun mit der Dichtefunktion aus der letzten Teilaufgabe eine Maximum-Likelihood-Schätzung durchführen. Im Vektor `x` sei dazu eine Stichprobe gespeichert.

10. Vervollständigen Sie die nachstehende Funktion, um sich, gegeben `x`, den log-Likelihood-Wert in Abhängigkeit von μ ausgeben lassen zu können:

```
logLik=function(mu,x){
                       g(    ,    )
}
```

Folgender `R`-Output zeigt die Funktion `logLik` für Werte von $\mu \in \{1, ..., 10\}$ gegeben `x`:



11. Identifizieren und markieren Sie auf der entsprechenden Achse der voranstehenden Grafik den ML-Schätzer.

Schmierpapier zu Aufgabe 2

Aufgabe 3

Teil I:

1. Eine Investition hat zum Zeitpunkt t_0 einen Wert von 100 EUR. In den darauf folgenden zehn Jahren beträgt die durchschnittliche Wachstumsrate 3.2 Prozent. Welchen Wert hat die Investition nach Ablauf der zehn Jahre?

2. Sie beobachten folgende Preisentwicklung einer Aktie A über fünf Jahre. Angegeben ist jeweils der Preis in EUR zum Jahresende.

Jahr i	1	2	3	4	5
Preis Aktie A in EUR	100	90	130	150	140

   Ein Anleger kauft am Ende von Jahr 1 insgesamt zehn Aktien. Wie hoch ist seine durchschnittliche Rendite pro Jahr am Ende von Jahr 5?

3. In drei aufeinanderfolgenden Jahren kauft ein Anleger Anteile einer Aktie. Er investiert 800 EUR im ersten Jahr, 600 EUR im zweiten Jahr und 200 EUR im dritten Jahr. Der Preis pro Anteil beträgt 25 EUR im ersten Jahr und 30 EUR im zweiten Jahr. Ferner ist über die drei Jahre ein durchschnittlicher Preis pro Anteil von 28.07 EUR bezahlt worden. Wie hoch ist der Preis pro Anteil im dritten Jahr?

Teil II:

4. An den Finanzmärkten treten immer wieder Extremereignisse auf, also besonders hohe oder niedrige Renditen. Ihnen sei folgende Zufallsvariable gegeben:

   X: „Anzahl der eingetretenen Extremereignisse pro Jahr“

   Ein Börsenjahr bestehe aus 250 Tagen. Es ist bekannt, dass im Mittel zwei Extremereignisse pro Jahr auftreten. Nehmen Sie an, dass X Poisson-verteilt ist.

   a) Nennen Sie die drei Annahmen des Poisson-Prozesses (nur verbal).

   b) Welcher Verteilung folgt die Wartezeit W zwischen zwei Poisson-Ereignissen?

   c) Berechnen Sie die Fläche unter der Dichtefunktion der Exponentialverteilung mit Parameter λ allgemein bis zu einem Punkt $x \geq 0$ (Rechenweg erforderlich).

   d) Wie hoch ist die Wahrscheinlichkeit, dass die Wartezeit W zwischen zwei Extremereignissen mehr als zwei Jahre (= 500 Tage) beträgt?

Teil III:

5. Ihnen sei die Zufallsvariable Y: „Rendite des Anlageprodukts Alpha in Prozent“ gegeben. Die Zufallsvariable Y folge einer t-Verteilung mit 5 Freiheitsgraden.

   a) Geben Sie den Erwartungswert von Y an.

   b) Bestimmen Sie die Varianz von Y.

   c) Welche Rendite wird mit einer Wahrscheinlichkeit von 99.9% nicht unterschritten?

   d) Ihnen sei weiterhin die Zufallsvariable Z: „Rendite des Anlageprodukts Beta in Prozent“ gegeben. Die Zufallsvariable Z folge einer Normalverteilung mit $\mu = 0$ und $\sigma = 1.29$. Welche Rendite wird nun mit einer Wahrscheinlichkeit von 99.9% nicht unterschritten?

Schmierpapier zu Aufgabe 3

Aufgabe 4

Sie sind Lehrer an einer Wirtschaftsschule und haben eine neue Methode entwickelt, Schülern das Fach Wirtschaft zu lehren. Sie wissen, dass bei Wirtschaftsschulaufgaben im Mittel 14 Punkte mit einer Standardabweichung von 7 Punkten erreicht werden. Die Ergebnisse einer Schulaufgabe die von 35 Schülern geschrieben wurde, die Sie unterrichtet haben, sehen Sie in der folgenden Tabelle:

Punkte	**Anzahl Schüler** n_i
0 bis unter 5	3
5 bis unter 10	3
10 bis unter 15	7
15 bis unter 20	12
20 bis unter 25	6
25 bis unter 30	4

1. Berechnen Sie den Mittelwert der erreichten Punkte in dieser Schulaufgabe.

Hinweis: Falls Sie diese Aufgabe nicht lösen konnten, nehmen Sie im Weiteren einen Mittelwert von 16.5 Punkten an.

Sie haben die Vermutung, dass Schüler, die mit Ihrer Methode unterrichtet wurden, bessere Ergebnisse in Schulaufgaben erzielen.

2. Um diese Vermutung zu überprüfen, testen Sie zuerst bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$, ob Ihre Daten normalverteilt sind. Dazu haben Sie sich folgende Tabelle als Arbeitshilfe (auf 4 Nachkomma-Stellen gerundet) erstellt:

Punkte	p_i^0	np_i^0
$-\infty$ bis unter 5	0.0980	3.4300
5 bis unter 10	0.1831	6.4085
10 bis unter 15	0.2729	
15 bis unter 20		8.6631
20 bis unter 25	0.1365	4.7985
25 bis $+\infty$	0.0620	2.1700

a) Vervollständigen Sie die Angaben in der Arbeitstabelle.

Sie fassen die erste und zweite Kategorie sowie die vorletzte und letzte zu je einer Kategorie zusammen.

b) Nennen Sie den Grund für dieses Vorgehen.

c) Bestimmen Sie die kritische Schranke für Ihren Test auf Vorliegen einer Normalverteilung bei Ihren Daten.

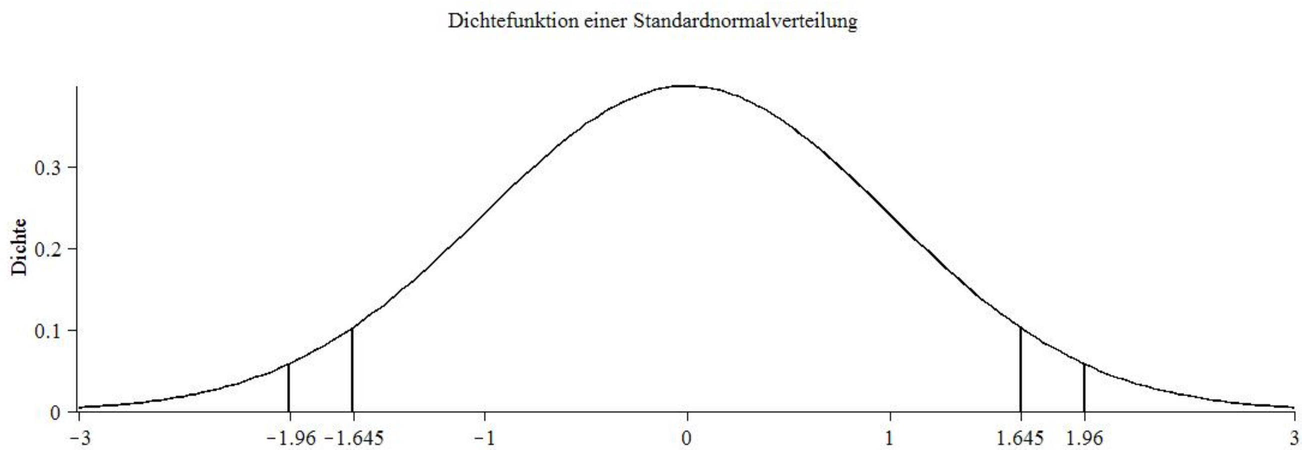
Arbeiten Sie im Weiteren mit der Normalverteilungsannahme für Ihre Daten. Die Standardabweichung der Grundgesamtheit gilt weiterhin als bekannt mit $\sigma = 7$.

3. Nun wollen Sie die Vermutung überprüfen, dass Ihre neue Lehrmethode signifikant höhere durchschnittliche Punktzahlen in Wirtschaftsschulaufgaben liefert. Somit ist:

$$H_0 : \mu \leqslant 14 \quad \text{gegen} \quad H_1 : \mu > 14$$

a) Erklären Sie die Bedeutung des Begriffs "signifikant".

b) Der folgende Plot zeigt die Dichtefunktion einer Standardnormalverteilung mit $\mu = 0$ und $\sigma^2 = 1$. Markieren Sie den Ablehnungsbereich mit der Irrtumswahrscheinlichkeit von $\alpha = 5\%$ für das Hypothesenpaar in diesem Plot eindeutig.



c) Berechnen Sie die realisierte Prüfgröße.

Falls Sie die letzte Aufgabe nicht lösen konnten, verwenden Sie als Wert der realisierten Prüfgröße: $t_n = 2.00$.

d) Berechnen Sie den p-Wert und treffen Sie eine Testentscheidung.

Nehmen Sie im Folgenden an, dass ein signifikanter Punktunterschied zwischen den Ergebnissen der Schüler, die Sie mit Ihrer neuen Lehrmethode unterrichteten und den allgemeinen Ergebnissen von Wirtschaftsschulaufgaben vorliegt.

4. Nun interessiert es Sie, ob dies ein großer oder vernachlässigbar kleiner Unterschied ist.

a) Berechnen Sie dazu den Schätzer der Effektgröße Cohen's δ:

Nehmen Sie an, Sie erhalten als Effektgröße $\delta = 0.3$.

b) Interpretieren Sie diesen Effekt anhand von Cohen's Interpretations-Grenzen Vorschlag.

5. Sie unterhalten sich mit dem Direktor der Schule über Ihre Ergebnisse. Dieser fragt Sie nach der Wahrscheinlichkeit des Fehlers 2. Art Ihres Tests.

a) Erklären Sie was der Fehler 2. Art aussagt.

Sie berechnen die Wahrscheinlichkeit des Fehlers 2. Art für den in 3. durchgeführten Hypothesentest und erhalten $\beta = 0.36$.

b) Berechnen Sie die Güte des Tests.

c) Nennen Sie eine Möglichkeit, wie die Güte des Tests erhöht werden kann.

d) Berechnen Sie den Stichprobenumfang, den Sie brauchen um eine Testentscheidung für den in 3. durchgeführten Hypothesentest mit einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ und einer Güte von 80% für $\mu_1 = 16.5$ treffen zu können.

Schmierpapier zu Aufgabe 4

Lösung Klausur SS15 (10ECTS)

Aufgabe 1

1. 0.28125. 0.5P

2. Modus G_1 Modalhäufigkeit 0.5625
   56,25% der Teilnehmer wurden der Gruppe der Spieler zugeordnet, die ein niedrigeres Handicap haben. 1.5P

3. $H_2 = 1 - 0.5625^2 - 0.4375^2 = 0.4922$
   Obergrenze: 1/2 hohe Streuung 2P

4. $n_{2,1} = 32 \cdot 0.125 = 4$
   4 Personen wurden der Gruppe 1 zugeordnet und benötigten drei Schläge. 1P

5. 15 Personen brauchten 5 Schläge. 0.5P

6. Nein, z.B. $0,09375 \cdot 0,5625 = f_{1,\cdot} \cdot f_{\cdot,1} \neq f_{1,1} = 0{,}0625$ 1.5P

7. $f(s_1|g_2) = \frac{f(s_1 \cap g_2)}{f(g_2)} = \frac{0{,}03125}{0{,}4375} = 0{,}0714$ 1P

8. Mit bedingten Häufigkeiten $f(S = S_i|G = G_1) = \frac{f_{i,1}}{f_{\cdot,1}}$:

S_i	S_1	S_2	S_3	S_4
$f(S = S_i\|G = G_1)$	0.1111111	0.2222222	0.2777778	0.3888889

$$\begin{aligned}\bar{s}_{|(G=G_1)} &= 0.1111111 * 2 + 0.2222222 * 3 + 0.2777778 * 4 + 0.3888889 * 5 \\ &= 3,9444\end{aligned}$$

2P

Aufgabe 2

Wegen `attach(schwangerschaft)` muss beim Zugriff auf den Data Frame kein `schwangerschaft$` verwendet werden

1. Nein, da `character` nur für qualitative Merkmale geeinget ist und *hCG-Wert* ein quantitatives Merkmal ist 0.5 P

2. z. B. `min(hCG.Wert)` 0.5 P (Quantil) + 0.5 P (Zugriff)

3. `hCG.Wert[Befund.Preggy=="positiv"|Befund.Baby=="positiv"]`
   1 P (-0.5 P pro Fehler)

4. `plot(ecdf(hCG.Wert))` 0.5 P (`plot`) + 0.5 P (`ecfd`)

5. (a) über 0.5 P
   (b) größer als 0.5 P

6. 0.4706 0.5 P

7. 34 0.5 P Wert + 0.5 P Wert ganzzahlig

8. 94.8529 0.5 P Berechnung + 0.5 P Wert

9. 0.5 P / 1 P (-0.5 P pro Fehler)
```
g=function(x,mu){
    sqrt(2*pi)^(-1)*exp(-(x-mu)^2)
}
```

10. 1 P (-0.5 P pro Fehler)
```
logLik=function(mu,x){
  sum(log(g(x,mu)))
}
```

11. Wert 6 hat den höchsten Likelohood Wert und ist somit der ML-Schätzer. 0.5 P

Aufgabe 3

1. 137.02 — 0.5 P+0.5 P
2. 0.0878 — 0.5 P
3. 40.00 — 0.5 P+0.5 P
4. a) P1 Wahrscheinlichkeit, in einem Intervall der Länge h das Ereignis genau einmal zu beobachten ist proportional zur Länge des Intervalls, aber unabhängig von dessen Lage. — 0.5 P

   P2 Wahrscheinlichkeit, in einem sehr kleinen Intervall der Länge h mehr als ein Ereignis zu beobachten ist vernachlässigbar klein. — 0.5 P

   P3 Die Anzahl der Ereignisse in zwei Intervallen, welche sich nicht überlappen, sind voneinander unabhängig. — 0.5 P

   b) Exponentialverteilt — 0.5 P

   c) $F_X(x; \lambda) = 1 - e^{-\lambda x}$ — 0.5 P+0.5 P

   d) 0.0183 — 0.5 P+0.5 P
5. a) 0 — 0.5 P

   b) 5/3 — 0.5 P+0.5 P

   c) -5.893 — 0.5 P+0.5 P

   d) -3.986 — 0.5 P+0.5 P

Aufgabe 4

1. Mittelwert = 16.3571 — 0.5 P

2. a) $p_4^0 = 0.2475$ — 0.5 P
   $np_3^0 = 9.5515$ — 0.5 P

   b) Bedingung für gute Approximation: $np_i^0 \geqslant 5$ für k $\leqslant 8$ — 0.5 P

   c) $\chi^2(k-1=3) = 7.81$ — 0.5 P

3. a) Signifikant = "nicht-zufällig" — 0.5 P

   b) nur Fläche auf der rechten Seite markiert — 0.5 P
   Richtige Schranke $\lambda_{0.95} = 1.6448(1.645)$ gewählt — 0.5 P

   c) $T_n = \sqrt{n}\frac{\overline{x}_n-\mu_0}{\sigma} = \sqrt{35}\frac{(16.3571-14)}{7} = 1.9921$
   $(\sqrt{35}\frac{(16.5-14)}{7} = 2.1129)$ — 0.5 P+0.5 P

   d) p-Wert $= P(T_n > 2.1129(1.9921)(2.00)) = 0.0174(0.0233)(0.0228)$ — 0.5 P
   $\Rightarrow$ Testentscheidung: Lehne Nullhypothese ab. — 0.5 P

4. a) Cohen's $\delta = \frac{16.3571-14}{7} = 0.3367$
   $(\frac{16.5-14}{7} = 0.3571)$ — 0.5 P

   b) Cohen's δ von 0.3 (0.3367) (0.3571) $\Longrightarrow$ kleiner Effekt — 0.5 P

5. a) Fehler 2. Art = Fehler die Nullhypothese nicht abzulehnen, — 0.5 P
   obwohl sie falsch ist — 0.5 P

   b) $1 - 0.36 = 0.64$ — 0.5 P

   c) - Erhöhung von α
   - Erhöhung des Stichprobenumfangs n
   (- Erhöhung der Effektgröße
   (Differenz zwischen μ_1 und μ_0 erhöhen, bzw Varianz verringern)) — 0.5 P

   d) $n = ((\lambda_{1-\alpha} - \lambda_\beta)\frac{\sigma}{\mu_1-\mu_0}))^2$
   $= ((1.6448 + 0.8416)\frac{7}{16.5-14})^2 = 48.4683(=49)$ — 0.5 P + 0.5 P