Klausur Statistik (7,5 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Donnerstag, 26.02.2015
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

Sie haben in einem Supermarkt 10 Orangen gekauft und diese gewogen.

1. Die folgende Tabelle zeigt die relative Häufigkeitsverteilung des Gewichts von 10 Orangen.

Gewicht in g	234	243	249	250	253	257
relative Häufigkeit	0.1	0.1	0.1	0.2	0.3	0.2

Bestimmen Sie die folgenden Größen:

a) Median

b) Unteres und oberes Quartil

c) Spannweite

d) Arithmetisches Mittel

e) Nehmen Sie an der Mittelwert sei 249. Treffen Sie eine Aussage zur Art der Schiefe der Gewichtsverteilung und begründen Sie diese.

Im Großhandel weiß man, dass die Wahrscheinlichkeit für eine Orange in schlechtem Zustand $P(s) = 30\%$ beträgt. Außerdem werden die Orangen mit einer Wahrscheinlichkeit von $P(I) = 80\%$ aus Italien und mit einer Wahrscheinlichkeit von $P(B) = 20\%$ aus Brasilien geliefert. Die Wahrscheinlichkeit, dass eine Orange, die aus Brasilien kommt, in schlechtem Zustand ist beträgt 0.35.

2. Überprüfen Sie, ob die Ereignisse, eine Orange in schlechtem Zustand und eine Orange aus Brasilien zu erhalten, stochastisch unabhängig sind.

Nun zieht ein Einzelhändler eine Stichprobe von 100 Orangen. Insgesamt hat er 27 Orangen in schlechtem Zustand gefunden. Nehmen Sie an, dass die jeweiligen Züge der Stichprobe voneinander unabhängig sind.

3. Berechnen Sie das realisierte, approximative 95%-Konfidenzintervall für die Erfolgswahrscheinlichkeit p, Orangen in schlechtem Zustand zu bekommen.

Sie ziehen nun selbst zufällig eine Orange aus einer Kiste mit 20 Orangen und überprüfen, ob diese in gutem oder schlechtem Zustand ist.

4. Wie nennt sich ein solches Experiment?

5. Die Stichprobe mit der Anzahl der Orangen in schlechtem Zustand in einer Kiste (X) folgt einer Binomialverteilung mit den Parametern $n = 20$ und $p = 0.3$.

   a) Mit welcher Wahrscheinlichkeit befinden sich in einer Kiste genau 10 Orangen in schlechtem Zustand?

   b) Insgesamt betrachten Sie 40 Orangen. Mit welcher Wahrscheinlichkeit sind genau 17 Orangen in schlechtem Zustand?

   c) Mit welcher Wahrscheinlichkeit sind in einer Kiste höchstens 18 Orangen in gutem Zustand?

   d) Geben Sie den Erwartungswert und die Varianz der Anzahl der Orangen in gutem Zustand bei einer Lieferung von 100 Orangen an.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Zur Wahl des Staatsoberhauptes schlagen die zwei größten Parteien jeweils einen sympathischen Kandidaten vor.
John A. möchte die Beliebtheit dieser Kandidaten in der Bevölkerung untersuchen. Dazu nutzt er Daten eines Forschungsinstituts, das 72 zufällig ausgewählte Personen befragte, wie sich deren jeweilige Sympathie auf die beiden Kandidaten aufteilt.

Befragter	Sympathieanteil Kandidat 1	Sympathieanteil Kandidat 2	Alter des Befragten in Jahren
i	v_i	w_i	u_i
1	0.1	0.9	60
2	0.5	0.5	19
3	0.4	0.6	53
4	0.7	0.3	39
⋮	⋮	⋮	⋮
72	0.6	0.4	24

Tabelle 1: Auszug aus der Urliste, die John A. für seine Untersuchungen heranzieht.

1. Wie alt ist Befragter 72?

2. Welchen Kandidaten bevorzugt Befragter 4?

3. Welchen Merkmalstyp hat das Merkmal „Alter des Befragten in Jahren“? Geben Sie ein geeignetes Streuungsmaß an.

John A. vermutet, dass Kandidat 1 bei den Befragten unter 50 Jahren beliebter ist, als bei denen ab 50 Jahren. Dies soll mittels eines Mittelwertdifferenzentests überprüft werden. Die Normalverteilungsannahme stört John A. bei der Analyse, u.a. weil Anteilswerte nur Werte im Intervall $[0, 1]$ annehmen können. Um dennoch den Test durchführen zu können, sollen die Daten transformiert werden. Dazu müssen zunächst die geometrischen Mittel der Sympathieanteile für jeden Befragten gebildet werden.

4. Berechnen Sie das geometrische Mittel der Sympathieanteile v_1 und w_1, die Befragter 1 angegeben hat.

Die Abbildungsvorschrift der zu verwendenden Transformation lautet:

$$(v_i, w_i) \mapsto \ln\left(\frac{v_i}{\sqrt{v_i w_i}}\right),$$

wobei $\sqrt{v_i w_i} = g(v_i, w_i)$ das geometrische Mittel von v_i und w_i ist.

5. Das geometrische Mittel der Sympathieanteile, die der Befragte 2 vergeben hat, beträgt $g(v_2, w_2) = 0.5$, das der Sympathieanteile, die der Befragte 3 vergeben hat $g(v_3, w_3) = 0.4898$. Transformieren Sie die Sympathieanteile der Befragten 2 und 3.

Die transformierten Daten teilt John A. in eine Teilstichprobe $x_1, \dots x_{44}$ mit den Befragten die unter 50 Jahre alt sind und eine weitere Teilstichprobe $y_1, \dots, y_{28}$ mit den Befragten ab 50 Jahren.

Verwenden Sie folgende Zwischenergebnisse:
$\overline{x}_{44} = 0.1115$, $\sum_{i=1}^{44}(x_i - \overline{x}_{44})^2 = 8.6091$, $\overline{y}_{28} = -0.2915$ und $\sum_{i=1}^{28}(y_i - \overline{y}_{28})^2 = 5.6079$.

6. Führen Sie einen Mittelwertdifferenzentest für unverbundene normalverteilte Stichproben bei unbekannten, aber homogenen Varianzen zum Signifikanzniveau $\alpha = 0.05$ durch. Das Hypothesenpaar lautet

$$\mathrm{H}_0: \ \mu_X \leq \mu_Y \quad vs. \quad \mathrm{H}_1: \ \mu_X > \mu_Y.$$

a) Geben Sie die kritische Schranke an.

Verwenden Sie im Folgenden 1.645 **als kritische Schranke, auch wenn Sie hier einen anderen Wert herausbekommen haben.**

b) Bestimmen Sie den Wert der realisierten Prüfgröße.

Verwenden Sie im Folgenden 3.211 **als Wert der realisierten Prüfgröße, auch wenn Sie hier einen anderen Wert herausbekommen haben.**

c) Treffen Sie die Testentscheidung basierend auf der kritischen Schranke und begründen Sie diese.

7. Bei der Wahl des Staatsoberhaupts teilen sich die Wähler in zwei Gruppen. Gruppe A steht der Partei nahe, die Kandidat 1 vorgeschlagen hat. Gruppe B steht der Partei nahe, die Kandidat 2 vorgeschlagen hat. Folgende Angaben sind bekannt: 57% der Wähler gehören zu Gruppe A. Von diesen stimmen 83% für Kandidat 1. Von Gruppe B stimmen 96% für Kandidat 2.

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

   b) Welcher Anteil der Wähler, von denen man weiß, dass sie Kandidat 2 gewählt haben, kommt aus Gruppe A?

   c) Welcher Kandidat bekommt die meisten Stimmen?

Schmierpapier zu Aufgabe 2

Aufgabe 3

Eine mittelständische Brauerei stellt Biere der Sorten

b_1: Rotbier b_2: Pils b_3: Helles b_4: Radler b_5: Weizen

her.

1. Die fünf Biersorten sind Ausprägungen eines Merkmals B. Welchen Merkmalstyp hat dieses Merkmal? Nennen Sie die maximal mögliche Skalierung.

2. Die Brauerei geht davon aus, dass ihre Biersorten mit den Wahrscheinlichkeiten p_i ($i = 1, 2, ..., 5$) gekauft werden. Konkret handelt es sich um folgende Wahrscheinlichkeiten:

	Rotbier(b_1)	Pils(b_2)	Helles(b_3)	Radler(b_4)	Weizen(b_5)
p_i	0.15	0.2	0.35	0.2	0.1

a) Geben Sie ein geeignetes Lagemaß für das Merkmal B an und bestimmen Sie dieses.

b) Ein geeignetes Streuungsmaß für das Merkmal B ist die Gini-Entropie.

$$H_2(p) = \sum_{i=1}^{k} p_i(1 - p_i)$$

Zeigen Sie, dass die Gini-Entropie auch wie folgt dargestellt werden kann:

$$H_2(p) = 1 - \sum_{i=1}^{k} p_i^2$$

c) Für welche Verteilung nimmt die Gini-Entropie ihren Maximalwert und für welche ihren Minimalwert an? Wie lauten diese beiden Werte?

Im Rahmen einer Markterhebung werden in der Stadt N ($j = 1$) insgesamt $n_1 = 780$ und in der Stadt M ($j = 2$) insgesamt $n_2 = 1350$ Personen zufällig ausgewählt und befragt, welche Biersorte sie bevorzugen (Mehrfachnennungen sind ausgeschlossen).
Dabei stellt sich folgendes Befragungsergebnis in den Stichproben ein:

	Rotbier(b_1)	Pils(b_2)	Helles(b_3)	Radler(b_4)	Weizen(b_5)
n_{1i}	195	78	273	78	156
n_{2i}	135	202	486	122	405

Sei hierbei n_{ji} (p_{ji}) die Häufigkeit (Wahrscheinlichkeit), mit der in der Stadt j die i-te Biersorte bevorzugt wird.

3. Geben Sie - ohne Herleitung - den Wert des Maximum-Likelihoodschätzers der Wahrscheinlichkeit p_{11} (dass eine befragte Person in N Rotbier bevorzugt) an.

Verwenden Sie für alle weiteren Rechnungen folgende Schätzwerte - auch wenn Sie andere Werte bestimmt haben:

	Rotbier(b_1)	Pils(b_2)	Helles(b_3)	Radler(b_4)	Weizen(b_5)
p_{1i}	0.26	0.09	0.37	0.09	0.19
p_{2i}	0.08	0.17	0.39	0.08	0.28

4. Wie groß ist die geschätzte Wahrscheinlichkeit, dass

   a) eine Person in N Pils bevorzugt?

   b) eine Person in M Rotbier oder Radler bevorzugt?

   c) eine Person aus N kommt und Weizen bevorzugt?

5. Geben Sie den Maximum-Likelihoodschätzer der Gini-Entropie sowie dessen Wert für die Stadt N an.

Verwenden Sie für alle weiteren Berechnungen die folgenden Schätzwerte - auch wenn Sie andere Werte bestimmt haben:

$H_2(\hat{p}_1) = 0.7632$ und $H_2(\hat{p}_2) = 0.7478$

6. Die Brauerei vermutet, dass in N die Streuung unter den bevorzugten Biersorten größer ist als in M. Sie will ihre Vermutung mit einem Hypothesentest auf dem 10%-Signifikanzniveau überprüfen. Dazu stellt sie folgende Hypothesen auf:

$$H_0: \ H_2(p_1) \leq H_2(p_2) \qquad\qquad H_1: \ H_2(p_1) > H_2(p_2)$$

Ihnen sei bekannt, dass die zu diesem Hypothesentest gehörige Prüfgröße asymptotisch standardnormalverteilt ist:

$$T_{NM} = \frac{H_2(\hat{p}_1) - H_2(\hat{p}_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \overset{asy}{\sim} \mathcal{N}(0,1)$$

mit $\hat{\sigma}_j^2 = \frac{4}{n_j}\left(\sum_{i=1}^{k}\left(\frac{N_{ji}}{n_j}\right)^3 - \left(\sum_{i=1}^{k}\left(\frac{N_{ji}}{n_j}\right)^2\right)^2\right)$ für $j \in \{1, 2\}$

a) Bestimmen Sie die realisierte Prüfgröße, wenn Sie wissen, dass $\hat{\sigma}_1^2 = 0.00073$ und $\hat{\sigma}_2^2 = 0.00048$.

b) Berechnen Sie den asymptotischen p-Wert.
Sollten Sie in Teilaufgabe 6a) kein Ergebnis erhalten haben, verwenden Sie als Wert der realisierten Prüfgröße $\mathbf{t_{NM} = 0.43}$

c) Treffen Sie die Testentscheidung und begründen Sie diese.
Sollten Sie in Teilaufgabe 6b) keinen p-Wert bestimmt haben, verwenden Sie $\mathbf{p = 0.35}$

Schmierpapier zu Aufgabe 3

Aufgabe 4

Teil I:

1. Sie beobachten nachstehende tägliche Renditen von Aktie A. *Zur Erinnerung:* Tägliche Renditen sind tägliche Wachstumsraten.

Tag i	1	2	3	4	5
Aktie A	-0.01	-0.01	0.02	0.03	0.00

   Nehmen Sie an, Sie hätten zum Beginn von Tag 1 insgesamt 100 EUR in Aktie A investiert. Welchen Wert hat Ihre Investition in Aktie A am Ende von Tag 5?

2. Sie investieren zu Beginn von Tag 1 je 100 EUR in Aktie A und Aktie B. Nach 100 Tagen beträgt der Wert Ihrer Investition in Aktie A und Aktie B insgesamt 300 EUR. Berechnen Sie die durchschnittliche Wachstumsrate pro Tag für die gesamte Investitionssumme.

3. Ein Anleger entscheidet sich für eine Investition in eine Aktie. In drei aufeinanderfolgenden Jahren kauft er Anteile dieser Aktie. Er investiert 100 EUR im ersten Jahr, 200 EUR im zweiten Jahr und 300 EUR im dritten Jahr. Der Preis pro Anteil beträgt 20 EUR im ersten Jahr, 25 EUR im zweiten Jahr und 50 EUR im dritten Jahr. Berechnen Sie den durchschnittlichen Preis pro Anteil für den gesamten Zeitraum.

Teil II:

4. Ein Anleger hätte gerne ein genaueres Verständnis bezüglich des Risikos seiner Anlage. Nehmen Sie an, dass die täglichen Renditen seiner Anlage normalverteilt sind mit $\mu = 0.0006$ und $\sigma^2 = 0.0005$.

   a) Mit welcher Wahrscheinlichkeit liegt die Tagesrendite über 2%?

   b) Welche Tagesrendite wird mit einer Wahrscheinlichkeit von 1% unterschritten?

Teil III:

5. Ihnen sei die normalverteilte Zufallsvariable Z: „Rendite eines Anlageprodukts" gegeben. Sie vermuten, dass die mittlere Rendite dieses Anlageprodukts größer ist als 0. Sie möchten Ihre Vermutung mit einem geeigneten Hypothesentest auf dem 5%-Niveau überprüfen. Dafür betrachten Sie die Renditen $Z_1, Z_2, ..., Z_{101}$ der letzten 101 Tage als Zufallsstichprobe. Sie stellen folgende Hypothesen auf:

   $$H_0 : \mu \leq \mu_0 = 0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0 = 0$$

   a) Bestimmen Sie die realisierte Prüfgröße. Für die Stichprobe haben Sie bereits ermittelt, dass $\bar{Z}_{101} = 0.000544$ und $S^2_{101} = 0.0004508$.

b) **Rechnen Sie jetzt mit einem Wert für die realisierte Prüfgröße von 0.254 weiter.** Bestimmen Sie den zugehörigen p-Wert.

c) **Arbeiten Sie jetzt mit einem p-Wert von 0.5 weiter.** Treffen Sie die Testentscheidung und begründen Sie diese.

Teil IV:

Im Folgenden seien drei Zufallsvariablen definiert:

X: „Tägliche Rendite von Aktie Alpha“, wobei $X \sim N(\mu_X, \sigma_X^2)$
Y: „Tägliche Rendite von Aktie Beta“, wobei $Y \sim N(\mu_Y, \sigma_Y^2)$
S: „Differenz der täglichen Renditen von Aktie Alpha und Aktie Beta“, also $S = X - Y$

Die Korrelation zwischen den Zufallsvariablen X und Y betrage ρ_{XY}.

6. Wie lautet der Erwartungswert der Zufallsvariablen S? Geben Sie Ihr Ergebnis in Abhängigkeit von μ_X, μ_Y, σ_X, σ_Y und ρ_{XY} an.

7. Wie lautet die Varianz der Zufallsvariablen S? Geben Sie Ihr Ergebnis in Abhängigkeit von μ_X, μ_Y, σ_X, σ_Y und ρ_{XY} an. Für welchen Wert von ρ_{XY} wird dieser Ausdruck minimal?

8. Sie vermuten, dass die Korrelation der täglichen Renditen der Aktien Alpha und Beta größer ist als 0.8. Sie möchten diese Vermutung mit einem geeigneten Hypothesentest auf dem 5%-Niveau überprüfen. Dafür betrachten Sie die Renditen $X_1, X_2, ..., X_{250}$ sowie $Y_1, Y_2, ..., Y_{250}$ der letzten 250 Tage als bivariate Zufallsstichprobe. Sie stellen folgende Hypothesen auf:

$$H_0 : \rho \leq \rho_0 = 0.8 \qquad \text{gegen} \qquad H_1 : \rho > \rho_0 = 0.8$$

Nehmen Sie an, dass die Zufallsvariablen X und Y bivariat normalverteilt sind.

a) Bestimmen Sie die realisierte Prüfgröße. Für die bivariate Stichprobe mit einem Umfang von 250 Tagen haben Sie bereits ermittelt, dass $R = 0.88$ ist.

b) Bestimmen Sie den Wert der kritischen Schranke.

c) **Rechnen Sie nun mit einer realisierten Prüfgröße von 4.0 weiter.** Treffen Sie die Testentscheidung und begründen Sie diese.

Schmierpapier zu Aufgabe 4

Lösung Klausur WS1415 (7.5 ECTS)

Aufgabe 1

1. a) Median: 250 — 0.5 P

   b) unteres und oberes Quartil: 249; 253 — 0.5 P+0.5 P

   c) Spannweite: $257 - 234 = 23$ — 0.5 P

   d) $\overline{X}_{10}$=249.9 — 0.5 P

   e) Linksschief, da Mittelwert (249) < Median (250) < Modus (253)
   oder: da Schiefemaß negativ — 0.5 P+0.5 P

2. $P(s) \cdot P(B) = 0.3 \cdot 0.2 = 0.06$
   $P(s|B) = \frac{P(s \cap B)}{P(B)}$
   $P(s \cap B) = P(s|B) \cdot P(B) = 0.35 \cdot 0.2 = 0.07$
   $P(s \cap B) \neq P(s) \cdot P(B)$
   oder: $P(s) = 0.3 \neq P(s|B) = 0.35$ — 0.5 P
   $\Rightarrow$ abhängig — 0.5 P

3. $[0.1830; 0.3570]$ — 0.5 P+0.5 P

4. Bernoulli-Experiment — 0.5 P

5. a) $f_{Binom}(x = 10; n = 20, p = 0.3) = 0.0308$ — 0.5 P

   b) $f_{Binom}(x = 17; n = 40, p = 0.3) = 0.0314$ — 0.5 P+0.5 P

   c) $1 - F_{Binom}(x = 1; n = 20, p = 0.3) = 1 - 0.0076 = 0.9924$ — 0.5 P+0.5 P+0.5 P

   d) $\mu = n \cdot p = 100 \cdot 0.7 = 70$ — 0.5 P
   $\sigma^2 = n \cdot p(1 - p) = 100 \cdot 0.7 \cdot 0.3 = 21$ — 0.5 P

Aufgabe 2

1. 24 Jahre — 0.5 P

2. Kandidat 1 — 0.5 P

3. Quantitatives Merkmal , Varianz — 0.5 P+0.5 P

4. 0.3 — 0.5 P

5. 0 und -0.2025 — 0.5 P+0.5 P

6. a) 1.667 — 0.5 P

   b) $\overline{s}^2 = 0.2031$, $\frac{0.1115+0.2915-0}{\sqrt{0.2031}\sqrt{1/44+1/28}} = 3.6990$ — 0.5 P+0.5 P+0.5 P

   c) $3.211 > 1.667 \Rightarrow$ Nullhypothese ablehnen — 0.5 P+0.5 P

7. a) — 0.5 P+0.5 P+0.5 P+0.5 P

	$K1$	$K2$	Σ
A	0.4731	0.0969	0.57
B	0.0172	0.4128	0.43
Σ	0.4903	0.5097	1

   b) $\frac{0.0969}{0.5097} = 0.1901$ — 0.5 P+0.5 P

   c) Kandidat 2 — 0.5 P

Aufgabe 3

1. ungeordnet kategorial (qualitativ) , Nominalskala — 0.5 P+0.5 P

2. a) Modus , Helles (B_3) — 0.5 P+0.5 P

   b) $H_2(p) = \sum_{i=1}^{k} p_i(1-p_i) = \sum_{i=1}^{k}(p_i - p_i^2) = \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} p_i^2 = 1 - \sum_{i=1}^{k} p_i^2$ — 0.5 P

   c) Minimalwert=0 für Einpunktverteilung
   Maximalwert=$\frac{k-1}{k}$ für Gleichverteilung — 0.5 P+0.5 P+0.5 P+0.5 P

3. $\hat{p}_{11} = 0.25$ — 0.5 P

4. a) $P_1(B_2) = 0.09$ — 0.5 P

   b) $P_2(B_1 \cup B_4) = 0.08 + 0.08 = 0.16$ — 0.5 P+0.5 P

   c) $P(B_5 \cap N) = 0.19 \cdot \frac{780}{780+1350} = 0.06958$ — 0.5 P

5. $\hat{H}_2(p_1) = H_2(\hat{p}_1) = 1 - \sum_{i=1}^{5}(p_{1i})^2 = 1 - 0.2568 = 0.7432$ — 0.5 P+0.5 P

6. a) $t_{NM} = \dfrac{0.7632 - 0.7478}{\sqrt{0.00073 + 0.00048}} \approx 0.4427$ — 0.5 P

   b) $P(T_{NM} > 0.4427) = 1 - \Phi(0.4427) = 1 - 0.67 = 0.33$ — 0.5 P+0.5 P
   $(P(T_{NM} > 0.43) = 1 - \Phi(0.43) = 1 - 0.6664 = 0.3336)$

   c) $p = 0.33 > 0.1 = \alpha$
   $\Longrightarrow$ Nullhypothese kann auf dem 10%-Niveau nicht abgelehnt werden. — 0.5 P

Aufgabe 4

1. 102.97 — 0.5 P+0.5 P

2. 0.0041 — 0.5 P

3. 31.58 — 0.5 P+0.5 P

4. a) 0.1922 — 0.5 P+0.5 P

   b) -0.0514 — 0.5 P+0.5 P

5. a) 0.2575 — 0.5 P

   b) 0.4000 — 0.5 P+0.5 P

   c) $0.5 > 0.05 = \alpha$, also kann H_0 auf dem 5% Signifikanzniveau nicht abgelehnt werden — 0.5 P

6. $\mathrm{E}[S] = \mu_x - \mu_y$ — 0.5 P

7. $\mathrm{VAR}[S_i] = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y$
   minimal für $\rho_{XY} = 1$ — 0.5 P+0.5 P

8. a) 4.3558 — 0.5 P+0.5 P

   b) 1.6448 — 0.5 P

   c) $4.0 > 1.6448$, also kann H_0 auf dem 5% Signifikanzniveau abgelehnt werden — 0.5 P

Klausur Statistik (10 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Donnerstag, 26.02.2015 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

Sie haben in einem Supermarkt 10 Orangen gekauft und diese gewogen.

1. Die folgende Tabelle zeigt die relative Häufigkeitsverteilung des Gewichts von 10 Orangen.

Gewicht in g	234	243	249	250	253	257
relative Häufigkeit	0.1	0.1	0.1	0.2	0.3	0.2

Bestimmen Sie die folgenden Größen:

a) Median

b) Unteres und oberes Quartil

c) Spannweite

d) Arithmetisches Mittel

e) Nehmen Sie an der Mittelwert sei 249. Treffen Sie eine Aussage zur Art der Schiefe der Gewichtsverteilung und begründen Sie diese.

Im Großhandel weiß man, dass die Wahrscheinlichkeit für eine Orange in schlechtem Zustand $P(s) = 30\%$ beträgt. Außerdem werden die Orangen mit einer Wahrscheinlichkeit von $P(I) = 80\%$ aus Italien und mit einer Wahrscheinlichkeit von $P(B) = 20\%$ aus Brasilien geliefert. Die Wahrscheinlichkeit, dass eine Orange, die aus Brasilien kommt, in schlechtem Zustand ist beträgt 0.35.

2. Überprüfen Sie, ob die Ereignisse, eine Orange in schlechtem Zustand und eine Orange aus Brasilien zu erhalten, stochastisch unabhängig sind.

Nun zieht ein Einzelhändler eine Stichprobe von 100 Orangen. Insgesamt hat er 27 Orangen in schlechtem Zustand gefunden. Nehmen Sie an, dass die jeweiligen Züge der Stichprobe voneinander unabhängig sind.

3. Berechnen Sie das realisierte, approximative 95%-Konfidenzintervall für die Erfolgswahrscheinlichkeit p, Orangen in schlechtem Zustand zu bekommen.

Sie ziehen nun selbst zufällig eine Orange aus einer Kiste mit 20 Orangen und überprüfen, ob diese in gutem oder schlechtem Zustand ist.

4. Wie nennt sich ein solches Experiment?

5. Die Stichprobe mit der Anzahl der Orangen in schlechtem Zustand in einer Kiste (X) folgt einer Binomialverteilung mit den Parametern $n = 20$ und $p = 0.3$.

   a) Mit welcher Wahrscheinlichkeit befinden sich in einer Kiste genau 10 Orangen in schlechtem Zustand?

   b) Insgesamt betrachten Sie 40 Orangen. Mit welcher Wahrscheinlichkeit sind genau 17 Orangen in schlechtem Zustand?

   c) Mit welcher Wahrscheinlichkeit sind in einer Kiste höchstens 18 Orangen in gutem Zustand?

   d) Geben Sie den Erwartungswert und die Varianz der Anzahl der Orangen in gutem Zustand bei einer Lieferung von 100 Orangen an.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Sie haben von einem Flughafenbetreiber folgenden, anonymisierten Datensatz (180 Flüge von Maschinen gleichen Typs auf der gleichen Strecke) mit folgenden Merkmalen erhalten:

Merkmal		**Spaltenname**
Fluggesellschaft:	Die den Flug durchführende Gesellschaft	`Fluggesellschaft`
Landungen:	Anzahl anderer Landungen am Flughafen in den 60 Minuten vor Landung des jeweiligen Fluges	`Landungen`
Erste Klasse:	Flug mit First Class Bereich	`FirstClass`
Verspätung:	Verspätung in Minuten, die der Flug bei Ankunft am Gate hatte	`Verspaetung`

Dieser Datensatz steht Ihnen in `R` als Data Frame `data` zur Verfügung. Arbeiten Sie zunächst mit folgendem `R`-Output:

```
'data.frame':        180 obs. of  4 variables:
 $ Fluggesellschaft: Factor w/ 3 levels "A","B","C": 3 3 2 2 3 1 3 2 1 1 ...
 $ Landungen       : num  38 76 12 52 95 3 26 27 52 80 ...
 $ FirstClass      : logi  TRUE FALSE FALSE FALSE TRUE TRUE ...
 $ Verspaetung     : num  13.1 20.6 4.1 27.4 21.7 ...
```

1. Mit welchem Befehl wurde der obige Output erzeugt?

2. Welcher Datentyp wäre für die Spalte `Fluggesellschaft` ebenfalls geeignet?

3. Geben Sie, basierend auf dem bisher verfügbaren `R`-Output, das kleinstmögliche Intervall für den Wertebereich der Merkmalsausprägungen von `Landungen` an.

4. Geben Sie den `R`-Befehl an, mit dem Sie lediglich die Verspätungen derjenigen Flüge der Fluggesellschaft `B` angezeigt bekommen, welche einen First Class Bereich hatten.

5. Geben Sie einen R-Befehl an, mit dem Sie den Prozentsatz aller Flüge, die über einen First Class Bereich verfügen, bestimmen.

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

```
> n=nrow(data)
> Ver.1=data[data$Fluggesellschaft=="A","Verspaetung"]
> Ver.2=data[data$Fluggesellschaft=="B","Verspaetung"]
> Ver.3=data[data$Fluggesellschaft=="C","Verspaetung"]
> addmargins(table(Fluggesellschaft,FirstClass)/n)

                FirstClass
Fluggesellschaft      FALSE      TRUE       Sum
             A    0.1388889 0.1944444 0.3333333
             B    0.2055556 0.1277778 0.3333333
             C    0.1777778 0.1555556 0.3333333
             Sum  0.5222222 0.4777778 1.0000000

> p=c(0.01,0.05,0.95,0.99)
> quantile(data$Verspaetung,p,type=1)

     1%      5%     95%     99%
 0.7011  1.7005 36.8804 46.9836

> quantile(data$Landungen,p)

    1%     5%    95%    99%
  2.79   5.00 132.45 197.42

> quantile(data$Landungen,p,type=1)

 1%  5% 95% 99%
  2   5 132 199

> c(mean(Ver.1),mean(Ver.2),mean(Ver.3))

[1] 19.42123 17.22547 18.07574
```

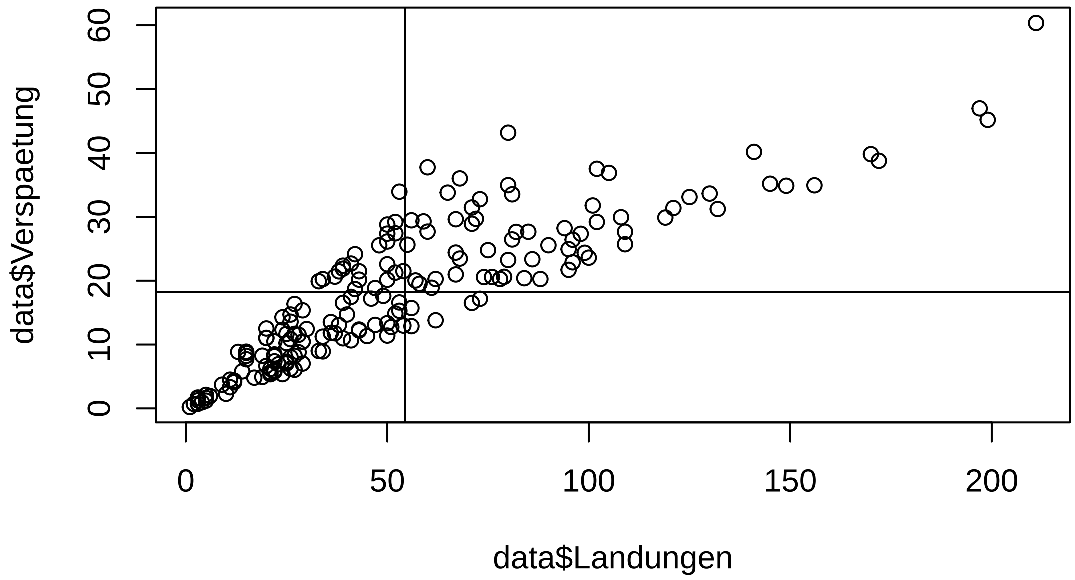
```
> table(data[data$FirstClass!=FALSE,"Fluggesellschaft"])

 A  B  C
35 23 28

> table(data[data$FirstClass!=TRUE,"Fluggesellschaft"])

 A  B  C
25 37 32
```

sowie die folgende mit `R` erstellte Grafik:



6. Welche Fluggesellschaft hat die meisten Flüge, welche über einen First Class Bereich verfügen? Wie viele Flüge waren das?

7. Vervollständigen Sie die folgenden Aussagen:

   a) Das Vorzeichen der Kovarianz zwischen den Merkmalen Landungen und Verspätung ist ________________.

   b) Es gab ________________ Flüge mit First Class Bereich als ohne.

   c) Fluggesellschaft ________________ war im Mittel pünktlicher als die anderen.

   d) Bei ________________% der Flüge landeten in der Stunde vor Landung höchstens 5 andere Flugzeuge.

8. Sie möchten die Korrelation aus den bereits vorhandenen Objekten `v1`, `v2` (Varianzen) und `Kova` (Kovarianz) berechnen. Vervollständigen Sie hierzu die folgende Funktion:

```
Kor=function(Kova,v1,v2){
          /(              *               )
}
```

Sie haben nun eine Stichprobe aus normalverteilter Grundgesamtheit `samp` in Ihrem Workspace.

9. Geben Sie den Befehl an, um den zu `samp` gehörigen Stichprobenumfang unter der Bezeichnung `n` abzuspeichern.

Im Folgenden seien unter `n` bzw. `S` Stichprobenumfang und -varianz von `samp` abgespeichert

10. Geben Sie den Code an, um die untere Schranke des realisierten, zweiseitigen 95%-Konfidenzintervalls des Mittelwerts der Grundgesamtheit zu berechnen (unbekannte Varianz). Verwenden Sie dabei **nicht** die Funktion `t.test()`!

Sie erhalten für untere bzw. obere Grenze des Konfidenzintervalls aus Teilaufgabe 10. folgenden `R`-Output:

```
[1] 12.41958 13.05348
```

11. Können Sie die Behauptung, der wahre Mittelwert der Grundgesamtheit würde sich vom Wert 13.05 unterscheiden, mit einer Irrtumswahrscheinlichkeit von 5% widerlegen? (Kurze Begründung)

Schmierpapier zu Aufgabe 2

Aufgabe 3

Eine mittelständische Brauerei stellt Biere der Sorten

b_1: Rotbier b_2: Pils b_3: Helles b_4: Radler b_5: Weizen

her.

1. Die fünf Biersorten sind Ausprägungen eines Merkmals B. Welchen Merkmalstyp hat dieses Merkmal? Nennen Sie die maximal mögliche Skalierung.

2. Die Brauerei geht davon aus, dass ihre Biersorten mit den Wahrscheinlichkeiten p_i $(i = 1, 2, ..., 5)$ gekauft werden. Konkret handelt es sich um folgende Wahrscheinlichkeiten:

	Rotbier(b_1)	Pils(b_2)	Helles(b_3)	Radler(b_4)	Weizen(b_5)
p_i	0.15	0.2	0.35	0.2	0.1

a) Geben Sie ein geeignetes Lagemaß für das Merkmal B an und bestimmen Sie dieses.

b) Ein geeignetes Streuungsmaß für das Merkmal B ist die Gini-Entropie.

$$H_2(p) = \sum_{i=1}^{k} p_i(1 - p_i)$$

Zeigen Sie, dass die Gini-Entropie auch wie folgt dargestellt werden kann:

$$H_2(p) = 1 - \sum_{i=1}^{k} p_i^2$$

c) Für welche Verteilung nimmt die Gini-Entropie ihren Maximalwert und für welche ihren Minimalwert an? Wie lauten diese beiden Werte?

Im Rahmen einer Markterhebung werden in der Stadt N ($j = 1$) insgesamt $n_1 = 780$ und in der Stadt M ($j = 2$) insgesamt $n_2 = 1350$ Personen zufällig ausgewählt und befragt, welche Biersorte sie bevorzugen (Mehrfachnennungen sind ausgeschlossen).
Dabei stellt sich folgendes Befragungsergebnis in den Stichproben ein:

	Rotbier(b_1)	Pils(b_2)	Helles(b_3)	Radler(b_4)	Weizen(b_5)
n_{1i}	195	78	273	78	156
n_{2i}	135	202	486	122	405

Sei hierbei n_{ji} (p_{ji}) die Häufigkeit (Wahrscheinlichkeit), mit der in der Stadt j die i-te Biersorte bevorzugt wird.

3. Geben Sie - ohne Herleitung - den Wert des Maximum-Likelihoodschätzers der Wahrscheinlichkeit p_{11} (dass eine befragte Person in N Rotbier bevorzugt) an.

Verwenden Sie für alle weiteren Rechnungen folgende Schätzwerte - auch wenn Sie andere Werte bestimmt haben:

	Rotbier(b_1)	Pils(b_2)	Helles(b_3)	Radler(b_4)	Weizen(b_5)
p_{1i}	0.26	0.09	0.37	0.09	0.19
p_{2i}	0.08	0.17	0.39	0.08	0.28

4. Wie groß ist die geschätzte Wahrscheinlichkeit, dass

   a) eine Person in N Pils bevorzugt?

   b) eine Person in M Rotbier oder Radler bevorzugt?

   c) eine Person aus N kommt und Weizen bevorzugt?

5. Geben Sie den Maximum-Likelihoodschätzer der Gini-Entropie sowie dessen Wert für die Stadt N an.

Verwenden Sie für alle weiteren Berechnungen die folgenden Schätzwerte - auch wenn Sie andere Werte bestimmt haben:

$H_2(\hat{p}_1) = 0.7632$ und $H_2(\hat{p}_2) = 0.7478$

6. Die Brauerei vermutet, dass in N die Streuung unter den bevorzugten Biersorten größer ist als in M. Sie will ihre Vermutung mit einem Hypothesentest auf dem 10%-Signifikanzniveau überprüfen. Dazu stellt sie folgende Hypothesen auf:

$$H_0: \; H_2(p_1) \leq H_2(p_2) \qquad\qquad H_1: \; H_2(p_1) > H_2(p_2)$$

Ihnen sei bekannt, dass die zu diesem Hypothesentest gehörige Prüfgröße asymptotisch standardnormalverteilt ist:

$$T_{NM} = \frac{H_2(\hat{p}_1) - H_2(\hat{p}_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \overset{asy}{\sim} \mathcal{N}(0,1)$$

mit $\hat{\sigma}_j^2 = \frac{4}{n_j}\left(\sum\limits_{i=1}^{k}\left(\frac{N_{ji}}{n_j}\right)^3 - \left(\sum\limits_{i=1}^{k}\left(\frac{N_{ji}}{n_j}\right)^2\right)^2\right)$ für $j \in \{1,\, 2\}$

a) Bestimmen Sie die realisierte Prüfgröße, wenn Sie wissen, dass $\hat{\sigma}_1^2 = 0.00073$ und $\hat{\sigma}_2^2 = 0.00048$.

b) Berechnen Sie den asymptotischen p-Wert.
Sollten Sie in Teilaufgabe 6a) kein Ergebnis erhalten haben, verwenden Sie als Wert der realisierten Prüfgröße $\mathbf{t_{NM} = 0.43}$

c) Treffen Sie die Testentscheidung und begründen Sie diese.
Sollten Sie in Teilaufgabe 6b) keinen p-Wert bestimmt haben, verwenden Sie $\mathbf{p = 0.35}$

Schmierpapier zu Aufgabe 3

Aufgabe 4

Teil I:

1. Sie beobachten nachstehende tägliche Renditen von Aktie A. *Zur Erinnerung:* Tägliche Renditen sind tägliche Wachstumsraten.

Tag i	1	2	3	4	5
Aktie A	-0.01	-0.01	0.02	0.03	0.00

Nehmen Sie an, Sie hätten zum Beginn von Tag 1 insgesamt 100 EUR in Aktie A investiert. Welchen Wert hat Ihre Investition in Aktie A am Ende von Tag 5?

2. Sie investieren zu Beginn von Tag 1 je 100 EUR in Aktie A und Aktie B. Nach 100 Tagen beträgt der Wert Ihrer Investition in Aktie A und Aktie B insgesamt 300 EUR. Berechnen Sie die durchschnittliche Wachstumsrate pro Tag für die gesamte Investitionssumme.

3. Ein Anleger entscheidet sich für eine Investition in eine Aktie. In drei aufeinanderfolgenden Jahren kauft er Anteile dieser Aktie. Er investiert 100 EUR im ersten Jahr, 200 EUR im zweiten Jahr und 300 EUR im dritten Jahr. Der Preis pro Anteil beträgt 20 EUR im ersten Jahr, 25 EUR im zweiten Jahr und 50 EUR im dritten Jahr. Berechnen Sie den durchschnittlichen Preis pro Anteil für den gesamten Zeitraum.

Teil II:

4. Ein Anleger hätte gerne ein genaueres Verständnis bezüglich des Risikos seiner Anlage. Nehmen Sie an, dass die täglichen Renditen seiner Anlage normalverteilt sind mit $\mu = 0.0006$ und $\sigma^2 = 0.0005$.

   a) Mit welcher Wahrscheinlichkeit liegt die Tagesrendite über 2%?

   b) Welche Tagesrendite wird mit einer Wahrscheinlichkeit von 1% unterschritten?

Teil III:

5. Ihnen sei die normalverteilte Zufallsvariable Z: „Rendite eines Anlageprodukts" gegeben. Sie vermuten, dass die mittlere Rendite dieses Anlageprodukts größer ist als 0. Sie möchten Ihre Vermutung mit einem geeigneten Hypothesentest auf dem 5%-Niveau überprüfen. Dafür betrachten Sie die Renditen $Z_1, Z_2, ..., Z_{101}$ der letzten 101 Tage als Zufallsstichprobe. Sie stellen folgende Hypothesen auf:

$$H_0 : \mu \leq \mu_0 = 0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0 = 0$$

   a) Bestimmen Sie die realisierte Prüfgröße. Für die Stichprobe haben Sie bereits ermittelt, dass $\bar{Z}_{101} = 0.000544$ und $S^2_{101} = 0.0004508$.

b) **Rechnen Sie jetzt mit einem Wert für die realisierte Prüfgröße von 0.254 weiter.** Bestimmen Sie den zugehörigen p-Wert.

c) **Arbeiten Sie jetzt mit einem p-Wert von 0.5 weiter.** Treffen Sie die Testentscheidung und begründen Sie diese.

Teil IV:

Im Folgenden seien drei Zufallsvariablen definiert:

X: „Tägliche Rendite von Aktie Alpha“, wobei $X \sim N(\mu_X, \sigma_X^2)$
Y: „Tägliche Rendite von Aktie Beta“, wobei $Y \sim N(\mu_Y, \sigma_Y^2)$
S: „Differenz der täglichen Renditen von Aktie Alpha und Aktie Beta“, also $S = X - Y$

Die Korrelation zwischen den Zufallsvariablen X und Y betrage ρ_{XY}.

6. Wie lautet der Erwartungswert der Zufallsvariablen S? Geben Sie Ihr Ergebnis in Abhängigkeit von μ_X, μ_Y, σ_X, σ_Y und ρ_{XY} an.

7. Wie lautet die Varianz der Zufallsvariablen S? Geben Sie Ihr Ergebnis in Abhängigkeit von μ_X, μ_Y, σ_X, σ_Y und ρ_{XY} an. Für welchen Wert von ρ_{XY} wird dieser Ausdruck minimal?

8. Sie vermuten, dass die Korrelation der täglichen Renditen der Aktien Alpha und Beta größer ist als 0.8. Sie möchten diese Vermutung mit einem geeigneten Hypothesentest auf dem 5%-Niveau überprüfen. Dafür betrachten Sie die Renditen $X_1, X_2, ..., X_{250}$ sowie $Y_1, Y_2, ..., Y_{250}$ der letzten 250 Tage als bivariate Zufallsstichprobe. Sie stellen folgende Hypothesen auf:

$$H_0 : \rho \leq \rho_0 = 0.8 \qquad \text{gegen} \qquad H_1 : \rho > \rho_0 = 0.8$$

Nehmen Sie an, dass die Zufallsvariablen X und Y bivariat normalverteilt sind.

a) Bestimmen Sie die realisierte Prüfgröße. Für die bivariate Stichprobe mit einem Umfang von 250 Tagen haben Sie bereits ermittelt, dass $R = 0.88$ ist.

b) Bestimmen Sie den Wert der kritischen Schranke.

c) **Rechnen Sie nun mit einer realisierten Prüfgröße von 4.0 weiter.** Treffen Sie die Testentscheidung und begründen Sie diese.

Schmierpapier zu Aufgabe 4

Lösung Klausur WS1415 (10 ECTS)

Aufgabe 1

1. a) Median: 250 — 0.5 P

   b) unteres und oberes Quartil: 249; 253 — 0.5 P+0.5 P

   c) Spannweite: $257 - 234 = 23$ — 0.5 P

   d) $\overline{X}_{10}$=249.9 — 0.5 P

   e) Linksschief, da Mittelwert (249) < Median (250) < Modus (253)
   oder: da Schiefemaß negativ — 0.5 P+0.5 P

2. $P(s) \cdot P(B) = 0.3 \cdot 0.2 = 0.06$
   $P(s|B) = \frac{P(s \cap B)}{P(B)}$
   $P(s \cap B) = P(s|B) \cdot P(B) = 0.35 \cdot 0.2 = 0.07$
   $P(s \cap B) \neq P(s) \cdot P(B)$
   oder: $P(s) = 0.3 \neq P(s|B) = 0.35$ — 0.5 P
   $\Rightarrow$ abhängig — 0.5 P

3. $[0.1830; 0.3570]$ — 0.5 P+0.5 P

4. Bernoulli-Experiment — 0.5 P

5. a) $f_{Binom}(x = 10; n = 20, p = 0.3) = 0.0308$ — 0.5 P

   b) $f_{Binom}(x = 17; n = 40, p = 0.3) = 0.0314$ — 0.5 P+0.5 P

   c) $1 - F_{Binom}(x = 1; n = 20, p = 0.3) = 1 - 0.0076 = 0.9924$ — 0.5 P+0.5 P+0.5 P

   d) $\mu = n \cdot p = 100 \cdot 0.7 = 70$ — 0.5 P
   $\sigma^2 = n \cdot p(1 - p) = 100 \cdot 0.7 \cdot 0.3 = 21$ — 0.5 P

Aufgabe 2

1. `str(data)` 0.5 P
2. Character 0.5 P
3. $[3; 95]$ 0.5 P
4. `data[data$Fluggesellschaft=="B" & data$FirstClass==TRUE,"Verspaetung"]`
   oder
   `data$Verspaetung[data$Fluggesellschaft=="B" & data$FirstClass==TRUE]`
   1 P (-0.5 P pro Fehler)
5. z. B. `sum(data$FirstClass==TRUE)/nrow(data)`
   0.5 P (Zähler) + 0.5 P (Nenner)
6. `A`, mit 35 Flügen. 0.5 P
7. a) positiv 0.5 P
   b) weniger 0.5 P
   c) `B` 0.5 P
   d) 5 0.5 P
8.
```
Kor=function(Kova,v1,v2){
    Kova/(sqrt(v1)*sqrt(v2))
}
```
   1 P (-0.5 P pro Fehler)
9. `n=length(samp)` 0.5 P
10. `mean(samp)-` 0.5 P
    `qt(0.975,df=n-1)` 0.5 P (t-Verteilung) + 0.5 P (Argumente)
    `*sqrt(S/n)` 0.5 P
11. Ja, weil 13.05 im realisierten 95%-Konfidenzintervall liegt 0.5 P

Aufgabe 3

1. ungeordnet kategorial (qualitativ) , Nominalskala 0.5 P+0.5 P

2. a) Modus , Helles (B_3) 0.5 P+0.5 P

   b) $H_2(p) = \sum_{i=1}^{k} p_i(1-p_i) = \sum_{i=1}^{k}(p_i - p_i^2) = \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} p_i^2 = 1 - \sum_{i=1}^{k} p_i^2$ 0.5 P

   c) Minimalwert=0 für Einpunktverteilung
      Maximalwert=$\frac{k-1}{k}$ für Gleichverteilung 0.5 P+0.5 P+0.5 P+0.5 P

3. $\hat{p}_{11} = 0.25$ 0.5 P

4. a) $P_1(B_2) = 0.09$ 0.5 P

   b) $P_2(B_1 \cup B_4)$= $0.08 + 0.08 = 0.16$ 0.5 P+0.5 P

   c) $P(B_5 \cap N) = 0.19 \cdot \frac{780}{780+1350} = 0.06958$ 0.5 P

5. $\hat{H}_2(p_1) = H_2(\hat{p}_1)$= $1 - \sum_{i=1}^{5}(p_{1i})^2 = 1 - 0.2568 = 0.7432$ 0.5 P+0.5 P

6. a) $t_{NM} = \dfrac{0.7632 - 0.7478}{\sqrt{0.00073 + 0.00048}} \approx 0.4427$ 0.5 P

   b) $P(T_{NM} > 0.4427)$= $1 - \Phi(0.4427) = 1 - 0.67 = 0.33$ 0.5 P+0.5 P
      $(P(T_{NM} > 0.43) = 1 - \Phi(0.43) = 1 - 0.6664 = 0.3336)$

   c) $p = 0.33 > 0.1 = \alpha$
      $\Longrightarrow$ Nullhypothese kann auf dem 10%-Niveau nicht abgelehnt werden. 0.5 P

Aufgabe 4

1. 102.97 — 0.5 P+0.5 P
2. 0.0041 — 0.5 P
3. 31.58 — 0.5 P+0.5 P
4. a) 0.1922 — 0.5 P+0.5 P

   b) -0.0514 — 0.5 P+0.5 P
5. a) 0.2575 — 0.5 P

   b) 0.4000 — 0.5 P+0.5 P

   c) $0.5 > 0.05 = \alpha$, also kann H_0 auf dem 5% Signifikanzniveau nicht abgelehnt werden — 0.5 P
6. $\mathrm{E}[S] = \mu_x - \mu_y$ — 0.5 P
7. $\mathrm{VAR}[S_i] = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y$

   minimal für $\rho_{XY} = 1$ — 0.5 P+0.5 P
8. a) 4.3558 — 0.5 P+0.5 P

   b) 1.6448 — 0.5 P

   c) $4.0 > 1.6448$, also kann H_0 auf dem 5% Signifikanzniveau abgelehnt werden — 0.5 P