Aufgabenstellung und Ergebnisse zur

Bachelor-Prüfung
Schliessende Statistik
Sommersemester 2021

PD Dr. Martin Becker

Hinweise für die Klausurteilnehmer

- Die Klausur besteht aus insgesamt 9 Aufgaben. Prüfen Sie die Vollständigkeit Ihres Exemplares nach; spätere Reklamationen können nicht berücksichtigt werden.
- Es sind insgesamt 120 Punkte (= 16 + 12 + 6 + 12 + 17 + 10 + 18 + 6 + 23) erreichbar.
- Als Hilfsmittel sind zugelassen: Taschenrechner (auch mit Grafikfähigkeit), 2 selbsterstellte DIN-A4 Blätter bzw. 4 selbsterstellte (einseitige) DIN-A4 Seiten. Benötigte Tabellen zur Normal- und t-Verteilung finden Sie am Ende dieses Klausurheftes.
- Bei der Korrektur werden **nur** die Lösungen auf den Seiten 1–11 berücksichtigt. Das letzte Blatt (Tabellen zur Normal- und t-Verteilung) darf abgetrennt werden.
- Bei mehreren Lösungsvorschlägen muss die gültige Lösung eindeutig gekennzeichnet sein.
- Mit Ausnahme der Multiple-Choice-Aufgaben muss der Lösungsweg klar ersichtlich sein.
- Alle Punkte innerhalb von Zahlen sind stets Dezimalpunkte (und keine Tausenderpunkte).

Bewertungsteil — Bitte nicht beschreiben							
Aufgabe	(a)	(b)	(c)	(d)	(e)	(f)	Σ
1							
2							
3							
4							
5							
6							
7							
8							
9							
Σ							

Aufgabe 1 (16 Punkte)

Markieren Sie jeweils mit einem Kreuz pro Aussage im betreffenden Kästchen, ob die unten stehenden Aussagen wahr oder falsch sind.
Richtige Antworten geben 2 Punkte, falsche Antworten und nicht bearbeitete Aussagen 0 Punkte (Aussagen mit zwei Kreuzen zählen als nicht bearbeitet!).

		wahr	falsch
1.	Stichproben zu normalverteilten Zufallsvariablen sind stets einfache Stichproben.	☐	■
2.	Sind für $n \in \mathbb{N}$ Schätzfunktionen T_n gegeben mit $\mathrm{E}(T_n) = \lambda$ für alle $n \in \mathbb{N}$, dann ist die Familie von Schätzfunktionen T_n stets konsistent im quadratischen Mittel für λ.	☐	■
3.	Bei Konfidenzintervallen für den Erwartungswert einer normalverteilten Zufallsvariablen mit unbekannter Varianz beeinflusst die Stichprobenrealisation nicht nur die Lage, sondern auch die Breite der realisierten Konfidenzintervalle.	■	☐
4.	Ist die Nullhypothese H_0 tatsächlich wahr, so wird man bei der Anwendung eines statistischen Tests zum Signifikanzniveau 0.01 nur mit einer Wahrscheinlichkeit von höchstens 1% eine Stichprobenrealisation erhalten, die zu einer Ablehnung von H_0 führt.	■	☐
5.	Entscheidet man sich bei der Durchführung eines statistischen Tests für H_0, obwohl H_0 tatsächlich nicht erfüllt ist, so begeht man einen Fehler 1. Art.	☐	■
6.	Der zum Vergleich von zwei (unbekannten) Erfolgswahrscheinlichkeiten p_A und p_B eingesetzte Test ist ein Spezialfall des (approximativen) 2-Stichproben-t-Tests zum Mittelwertvergleich, bei dem die dort üblicherweise vorauszusetzende Varianzgleichheit der beiden untersuchten Zufallsvariablen unter H_0 (im Fall $p_A = p_B$) automatisch gegeben ist.	■	☐
7.	Mit einem Chi-Quadrat-Unabhängigkeitstest soll anhand einer einfachen Stichprobe vom Umfang $n = 500$ die stochastische Unabhängigkeit zweier diskreter Zufallsvariablen mit jeweils 3 (auch in der Stichprobe mit zur Anwendung des Tests ausreichenden Häufigkeiten beobachteten) Trägerpunkten überprüft werden. Damit ist zur Konstruktion des Ablehnungsbereiches die χ^2-Verteilung mit 4 Freiheitsgraden zu verwenden.	■	☐
8.	Zur Parameterschätzung im einfachen linearen Regressionsmodell $$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2),$$ wird vorausgesetzt, dass die Werte x_i nicht alle übereinstimmen.	■	☐

Aufgabe 2 (12 Punkte)

Markieren Sie jeweils die korrekte Antwort mit einem Kreuz im betreffenden Kästchen. Es ist jeweils genau ein Kreuz korrekt.
Richtige Antworten geben 3 Punkte, falsche Antworten und nicht bearbeitete Aufgabenteile 0 Punkte (Aufgabenteile mit mehr als einem Kreuz zählen als nicht bearbeitet!).

1. Bei der Durchführung eines χ^2-Tests für die Varianz einer normalverteilten Zufallsvariablen bei unbekanntem Erwartungswert auf Grundlage einer einfachen Stichprobe vom Umfang n zum Signifikanzniveau 0.05 lehnt der linksseitige Test H_0 ab, während der zweiseitige Test H_0 nicht verwerfen kann. Damit weiß man über die Realisation χ^2 der Teststatistik:

   (a) $\chi^2 \in [0, \chi^2_{n-1;0.025})$ ☐

   (b) $\chi^2 \in [\chi^2_{n-1;0.025}, \chi^2_{n-1;0.05})$ ■

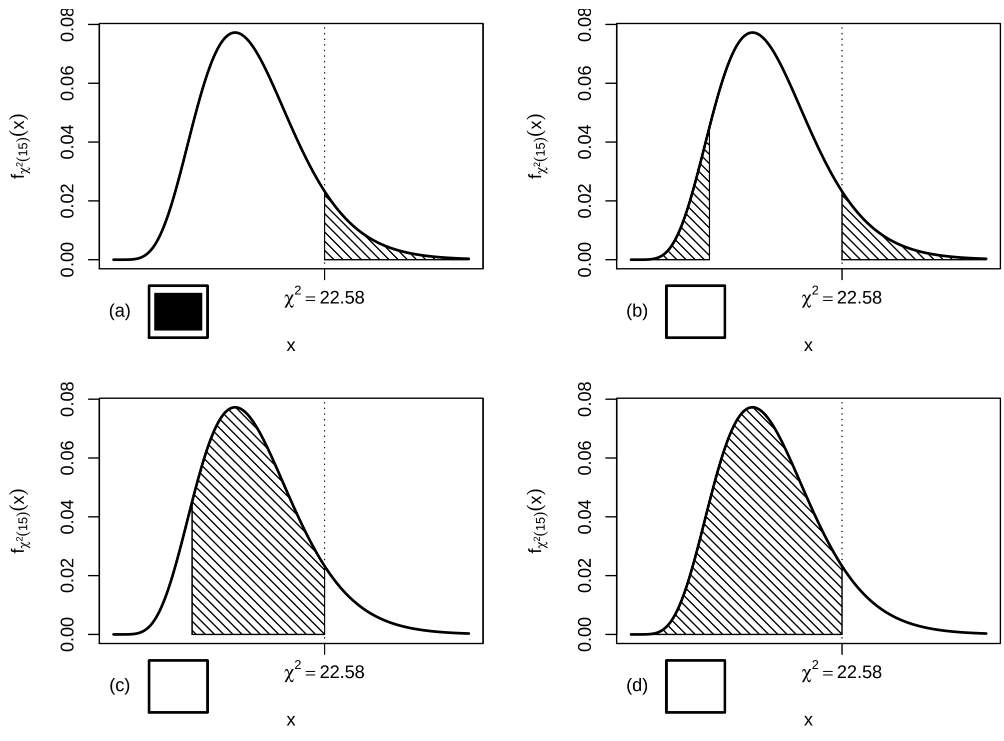
   (c) $\chi^2 \in [\chi^2_{n-1;0.05}, \chi^2_{n-1;0.95}]$ ☐

   (d) $\chi^2 \in [\chi^2_{n-1;0.05}, \infty)$ ☐

2. Sei $X_1, \dots, X_{16}$ eine einfache Stichprobe zu einer $N(\mu, \sigma^2)$-verteilten Zufallsvariablen Y mit unbekannten Parametern μ und σ^2. Auf der Grundlage einer Stichprobenrealisation zu dieser einfachen Stichprobe vom Umfang $n = 16$ soll

$$H_0 : \sigma^2 \le \sigma_0^2 = 9 \qquad \text{gegen} \qquad H_1 : \sigma^2 > \sigma_0^2 = 9$$

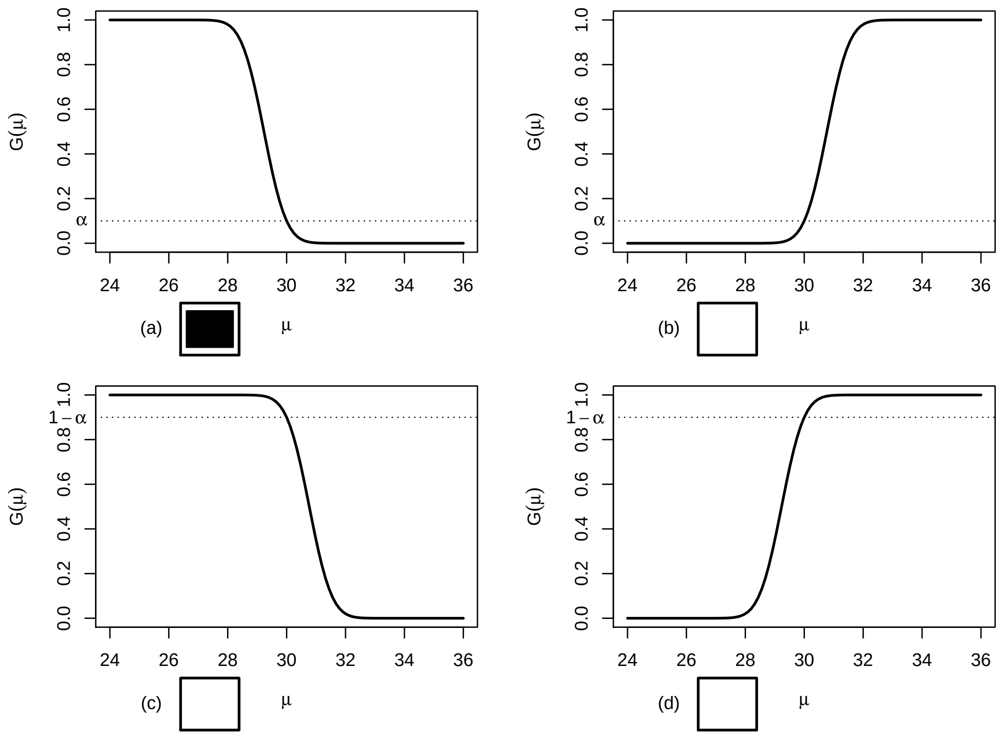
   mit einem Chi-Quadrat-Test getestet werden. Als realisierte Teststatistik erhält man $\chi^2 = 22.58$. Markieren Sie die Abbildung, welche den p-Wert in der beschriebenen Situation korrekt als Inhalt der schraffierten Fläche unter der Dichtefunktion der Verteilung der Teststatistik unter H_0 (für $\sigma^2 = \sigma_0^2$) darstellt.



3. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{25}$ vom Umfang $n = 25$ zu einer $N(\mu, 3^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu \geq 30 \qquad \text{gegen} \qquad H_1 : \mu < 30$$

bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.
Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.



4. Für die Erfüllung der angestrebten Güteeigenschaften von statistischen Tests ist es vorteilhaft, wenn
   (a) die Verteilung der Teststatistik bei Gültigkeit von H_0 exakt mit der Verteilung der Teststatistik bei Gültigkeit von H_1 übereinstimmt.
   (b) die Verteilung der Teststatistik bei Gültigkeit von H_0 wenigstens für große Stichprobenumfänge näherungsweise mit der Verteilung der Teststatistik bei Gültigkeit von H_1 übereinstimmt.
   (c) sich die Verteilungen der Teststatistik bei Gültigkeit von H_0 möglichst deutlich von den Verteilungen der Teststatistik bei Gültigkeit von H_1 unterscheiden.
   (d) sich die Verteilungen der Teststatistik bei Gültigkeit von H_0 möglichst wenig von den Verteilungen der Teststatistik bei Gültigkeit von H_1 unterscheiden.

Aufgabe 3 (3 + 3 = 6 Punkte)

In Abhängigkeit eines unbekannten Parameters p mit $0 < p < 1$ sei die Verteilung einer Zufallsvariablen Y gegeben durch:

y_i	-2	1	3
$p_Y(y_i)$	$\frac{1-p}{2}$	p	$\frac{1-p}{2}$

Für $n \in \mathbb{N}$ sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y.

(a) Zeigen Sie, dass $\mathrm{E}(Y) = \dfrac{1}{2} + \dfrac{1}{2} \cdot p$ gilt.

(b) Zeigen Sie, dass die Schätzfunktionen

$$T_n(X_1, \ldots, X_n) := 2 \cdot \overline{X} - 1$$

(mit $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$) für $n \in \mathbb{N}$ erwartungstreu für p sind.

Ergebnisse (ohne Begründung/Rechenweg):

(a) Berechnung des Erwartungswerts von Y.

(b) Berechnung des Erwartungswerts von $T_n(X_1, \ldots, X_n)$.

Aufgabe 4 (6 + 4 + 2 = 12 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $b > 2$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|b) = \begin{cases} \dfrac{2(y-2)}{(b-2)^2} & \text{für } 2 \leq y \leq b \\ 0 & \text{sonst} \end{cases}$$

Der Parameter b soll auf Grundlage einer einfachen Stichprobe $X_1, \ldots, X_n$ vom Umfang n geschätzt werden.

(a) Bestimmen Sie den Schätzer $\widehat{b}_{ML}$ nach der Maximum-Likelihood-Methode.

(b) Zeigen Sie, dass $E(Y) = \frac{2}{3}b + \frac{2}{3}$ gilt.

(c) Bestimmen Sie den Schätzer $\widehat{b}_{MM}$ nach der Methode der Momente.

Hinweise:

- *Beachten Sie, dass Sie Teil (b) auch ohne die Bearbeitung von Teil (a) und Teil (c) ohne die Bearbeitung der Teile (a) und (b) lösen können.*
- *Falls sich der ML-Schätzer als lokale Extremstelle einer differenzierbaren Funktion bestimmen lässt, muss* **nicht** *überprüft werden (z.B. mit Hilfe der 2. Ableitung), ob tatsächlich eine Maximalstelle vorliegt.*
- *Am Ende von Aufgabenteil (b) ist eine Polynomdivision oder eine Erweiterung der rechten Seite eventuell hilfreich.*

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{b}_{ML} = \max\{x_1, \ldots, x_n\}$

(b) Nachweis durch Berechnung des Erwartungswerts

(c) $\widehat{b}_{MM} = \frac{3}{2}\overline{x} - 1$

Aufgabe 5 $(7 + 2 + 4 + 4 = 17$ Punkte)

Eine Maschine produziert Schrauben, deren Zugfestigkeit erfahrungsgemäß normalverteilt mit einer Standardabweichung von $10[N/mm^2]$ um den tatsächlichen Erwartungswert schwankt. Die laufende Qualitätskontrolle soll eine Unterschreitung dieses Erwartungswerts gegenüber der mittleren Soll-Zugfestigkeit $800[N/mm^2]$ mit Hilfe eines geeigneten statistischen Testverfahrens auf Basis der Realisation einer einfachen Stichprobe $x_1, \ldots, x_{25}$ aufdecken. Dabei darf eine derartige Unterschreitung nur mit einer Wahrscheinlichkeit von höchstens 5% fälschlicherweise signalisiert werden. Aus dem realisierten Stichprobenergebnis erhält man den Stichprobenmittelwert

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 796.09[N/mm^2] .$$

(a) Führen Sie den zur oben beschriebenen Qualitätskontrolle geeigneten Test auf Basis des angegebenen Stichprobenmittelwerts durch. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Berechnen Sie den p-Wert zum Test aus Teil (a). Wie wäre die Entscheidung zum Test aus Teil (a) bei einem Signifikanzniveau von $\alpha = 0.01$ ausgefallen?

(c) Mit welcher Wahrscheinlichkeit wird die Qualitätskontrolle bei Ziehung einer einfachen Stichprobe der Länge 25 keine Unterschreitung signalisieren, wenn der tatsächliche Erwartungswert der Zugfestigkeit der Schrauben $794[N/mm^2]$ beträgt?

(d) Wie groß muss der Stichprobenumfang bei der oben beschriebenen Qualitätskontrolle mindestens gewählt werden, wenn eine tatsächliche Unterschreitung des Erwartungswerts der Zugfestigkeit um $6[N/mm^2]$ mit einer Wahrscheinlichkeit von mindestens 99% erkannt werden soll?

Ergebnisse (ohne Begründung/Rechenweg):

(a) $N = -1.95 \in (-\infty, -1.645) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

Der Test zur Qualitätskontrolle wird also eine Unterschreitung der mittleren Zugfestigkeit signalisieren.

(b) p-Wert $p = 0.025$. Der Test zur Qualitätskontrolle hätte also keine Unterschreitung der mittleren Zugfestigkeit signalisiert.

(c) $\beta(794) = 0.0869$

(d) Der Stichprobenumfang muss mindestens $n = 44$ betragen.

Aufgabe 6 (10 Punkte)

Die Leistungsfähigkeit zweier CPUs ("A" und "B") soll mit Hilfe von Benchmarks zur Leistungsmessung verglichen werden. Man nehme hierzu an, dass die erhaltenen Werte Y^A bzw. Y^B der Benchmarks zu CPU A bzw. CPU B jeweils normalverteilt seien mit den unbekannten Erwartungswerten μ_A bzw. μ_B sowie den unbekannten Varianzen σ_A^2 bzw. σ_B^2. Es soll überprüft werden, ob CPU B im Mittel höhere Benchmarkergebnisse als CPU A liefert.
Aus einer wiederholten Durchführung mit $n_A = 15$ Benchmark-Durchläufen für CPU A sowie $n_B = 13$ Durchläufen für CPU B erhielt man Realisationen jeweils voneinander unabhängiger einfacher Stichproben $X_1^A, \ldots, X_{15}^A$ zu Y^A sowie $X_1^B, \ldots, X_{13}^B$ zu Y^B und hieraus die zugehörigen Mittelwerte $\overline{x^A} = 6834$ bzw. $\overline{x^B} = 6895$ sowie die Stichproben**standardabweichungen** $s_{Y^A} = 120$ bzw. $s_{Y^B} = 89$. Testen Sie unter der Annahme $\sigma_A^2 = \sigma_B^2$ zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass CPU B im Mittel höhere Benchmarkergebnisse als CPU A liefert. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Ergebnisse (ohne Begründung/Rechenweg):

$t = -1.507 \notin (-\infty, -1.706) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
Der Test kann also die Vermutung, dass CPU B im Mittel höhere Benchmarkergebnisse als CPU A liefert, nicht bestätigen.

Aufgabe 7 (14 + 4 = 18 Punkte)

Um zu überprüfen, ob die Anzahl der bearbeiteten Zusatzübungsblätter einen Einfluss auf die Leistung in der schriftlichen Prüfung hat, wurden die Prüflinge einer Statistik-Klausur im Wintersemester 2020/21 in 3 Gruppen eingeteilt (Gruppe 1: kein Zusatzblatt bearbeitet, Gruppe 2: ein Zusatzblatt bearbeitet, Gruppe 3: beide Zusatzblätter bearbeitet). Aus den erreichten Punktzahlen der entsprechenden Klausur wurden auf Basis dieser Einteilung die folgenden Daten aggregiert:

j (Gruppe)	n_j	$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{j,i}$	$\sum_{i=1}^{n_j} x_{j,i}^2$	s_j^2
1	67	75.49	415902	516.46
2	18	90.17	158314	703.69
3	21	97.40	202187	148.25

(a) Überprüfen Sie mit einer einfachen Varianzanalyse (Signifikanzniveau $\alpha = 0.05$) unter der Annahme, dass die erreichten Punktzahlen Realisierungen von unabhängig $N(\mu_j, \sigma^2)$-verteilten Zufallsvariablen $X_{j,i} (1 \leq j \leq 3, 1 \leq i \leq n_j)$ sind, ob die Anzahl der bearbeiteten Zusatzblätter einen Einfluss auf die (mittlere) erreichte Punktzahl hat. Fassen Sie das Ergebnis der Varianzanalyse auch in einem Antwortsatz zusammen.

(b) Erläutern Sie im Hinblick auf den Unterschied zwischen s_2^2 und s_3^2 mit kurzer Begründung (1–2 Sätze), inwiefern die Gültigkeit der zur Anwendung der Variananalyse getroffenen Annahme der Varianzgleichheit in den zugehörigen Gruppen zu hinterfragen ist. *(Hinweis: $F_{17,20;0.025} = 0.382$, $F_{17,20;0.975} = 2.523$)*

Hinweis: Verwenden Sie für Teil (a) den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m,n)$*-Verteilungen:*

$n\backslash m$	1	2	3	4	5	100	101	102	103	104
1	161.448	199.500	215.707	224.583	230.162	253.041	253.054	253.066	253.078	253.090
2	18.513	19.000	19.164	19.247	19.296	19.486	19.486	19.486	19.486	19.486
3	10.128	9.552	9.277	9.117	9.013	8.554	8.554	8.553	8.553	8.553
4	7.709	6.944	6.591	6.388	6.256	5.664	5.664	5.663	5.663	5.663
5	6.608	5.786	5.409	5.192	5.050	4.405	4.405	4.404	4.404	4.404
100	3.936	3.087	2.696	2.463	2.305	1.392	1.391	1.390	1.389	1.388
101	3.935	3.086	2.695	2.462	2.304	1.390	1.389	1.388	1.388	1.387
102	3.934	3.085	2.694	2.461	2.303	1.389	1.388	1.387	1.386	1.385
103	3.933	3.085	2.693	2.460	2.303	1.388	1.387	1.386	1.385	1.384
104	3.932	3.084	2.692	2.459	2.302	1.386	1.385	1.385	1.384	1.383

Ergebnisse (ohne Begründung/Rechenweg):

(a) $F = 9.467 \in (3.085, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

Die einfache Varianzanalyse kommt also zum Ergebnis, dass die Anzahl der bearbeiteten Zusatzblätter einen signifikanten ($\alpha = 0.05$) Einfluss auf die (mittlere) erreichte Punktzahl hat.

(b) Es gibt Anzeichen für eine Verletzung der angenommenen Varianzgleichheit.

Aufgabe 8 (1 + 1 + 1 + 1 + 1 + 1 = 6 Punkte)

Zur Erklärung der stetigen Wochenrenditen der Siemens-Aktie y_i (in Prozent) durch die stetigen Wochenrenditen des DAX x_i (in Prozent) unterstellt man die Gültigkeit eines Zusammenhangs im Sinne des folgenden linearen Modells:

$$y_i = \beta_1 + \beta_2 x_i + u_i \qquad \text{mit} \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2),\ i \in \{1, \ldots, n\}$$

Aus Daten der XETRA-Börse wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5422 -0.5705 -0.0089  0.5355  2.5881

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1813     0.4153  -0.437   0.6690
x             0.5344     0.2901   1.842   0.0867 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.491 on 14 degrees of freedom
Multiple R-squared:  0.1951,       Adjusted R-squared:  0.1376
F-statistic: 3.394 on 1 and 14 DF,  p-value: 0.08671
```

(a) Geben Sie die realisierten Kleinst-Quadrate-Schätzwerte für β_1 und β_2 an.

(b) Geben Sie den realisierten Schätzwert für σ^2 an.

(c) Welcher Anteil der Gesamtvarianz der stetigen Wochenrenditen der Siemens-Aktie wird durch das lineare Modell erklärt?

(d) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.10$, ob β_1 signifikant von Null verschieden ist.

(e) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.05$, ob β_2 signifikant positiv ist.

(f) Welche stetige Wochenrendite der Siemens-Aktie prognostiziert das Modell in einer Woche mit stetiger DAX-Rendite von 0.4 (in Prozent)?

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_1 = -0.1813$, $\widehat{\beta}_2 = 0.5344$

(b) $\widehat{\sigma^2} = 2.2231$

(c) 0.1951

(d) β_1 ist nicht signifikant von Null verschieden.

(e) β_2 ist signifikant positiv.

(f) 0.03246

Aufgabe 9 (6 + 2 + 2 + 3 + 5 + 5 = 23 Punkte)

Zur Schätzung eines einfachen linearen Regressionsmodells

$$y_i = \beta_1 + \beta_2 \cdot x_i + u_i \qquad \text{mit} \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad i \in \{1, \ldots, n\}$$

aus einer Stichprobe vom Umfang $n = 30$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{30} y_i = 447.287; \quad \sum_{i=1}^{30} y_i^2 = 8665.384; \quad \sum_{i=1}^{30} x_i = 202.524;$$

$$\sum_{i=1}^{30} x_i^2 = 1537.306; \quad \sum_{i=1}^{30} x_i \cdot y_i = 3578.792$$

(a) Schätzen Sie β_1 und β_2 mit Hilfe der Kleinst-Quadrate-Methode.

(b) Berechnen Sie das Bestimmtheitsmaß R^2.

(c) Geben Sie mit Hilfe der bekannten erwartungstreuen Schätzfunktion für σ^2 den realisierten Schätzwert für σ^2 an.

(d) Berechnen Sie $\widehat{\sigma^2}_{\widehat{\beta}_1}$ und $\widehat{\sigma^2}_{\widehat{\beta}_2}$.

(e) Überprüfen Sie zum Signifikanzniveau $\alpha = 0.01$ **(!)**, ob β_1 signifikant negativ ist. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

(f) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für y_0 gegeben $x_0 = 5$ an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_1 = -7.2843$, $\widehat{\beta}_2 = 3.2876$

(b) $R^2 = 0.9209$

(c) $\widehat{\sigma}^2 = 5.6405$

(d) $\widehat{\sigma^2}_{\widehat{\beta}_1} = 1.6992$, $\widehat{\sigma^2}_{\widehat{\beta}_2} = 0.033159$

(e) $t = -5.588 \in (-\infty, -2.467) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

β_1 ist also signifikant negativ.

(f) $[4.167, 14.141]$