Aufgabenstellung und Ergebnisse zur

Bachelor-Prüfung
Schliessende Statistik
Sommersemester 2018

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 + 11 + 13 + 20 + 15 + 6 + 21) 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!).

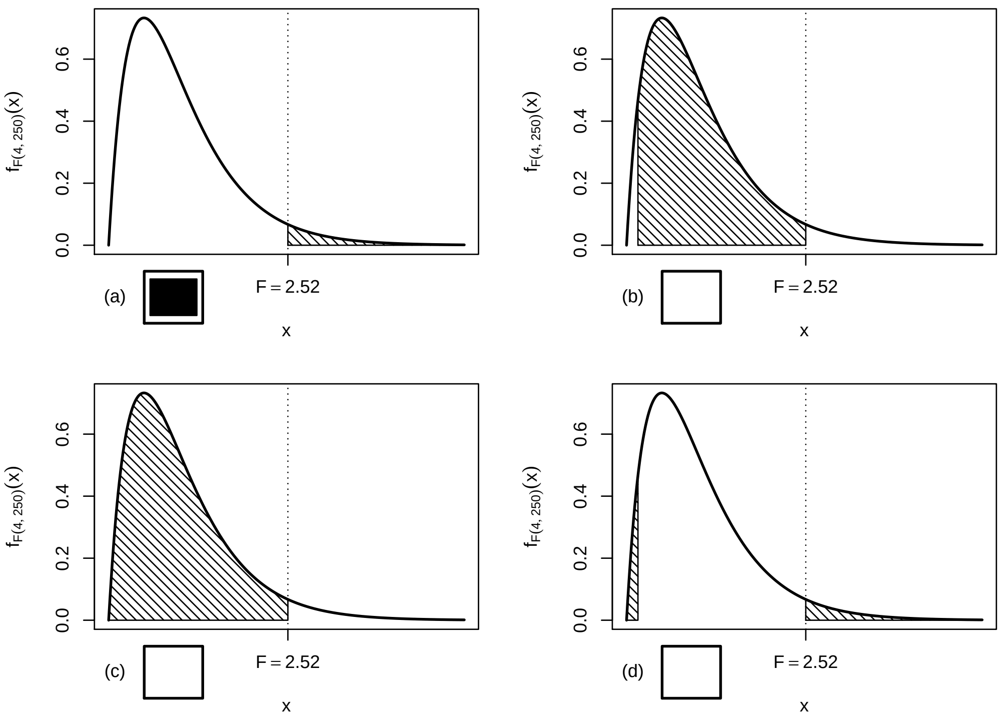
		wahr	falsch
1.	Sei $X_1, \dots, X_n$ eine einfache Stichprobe vom Umfang n zu Y. Dann sind $X_1, \dots, X_n$ stets normalverteilt.	☐	■
2.	Sind für $n \in \mathbb{N}$ Schätzfunktionen T_n gegeben, für die $\mathrm{E}(T_n) = \lambda$ sowie $\lim_{n\to\infty} \mathrm{Var}(T_n) = 0$ gilt, dann ist die Familie von Schätzfunktionen T_n stets effizient für λ.	☐	■
3.	Zur Schätzung des Parameters $\theta \in \mathbb{R}$ seien für $n \in \mathbb{N}$ Schätzfunktionen T_n gegeben mit den Eigenschaften $\mathrm{E}(T_n) = \theta + \frac{1}{n}$ und $\mathrm{Var}(T_n) = \frac{16}{n}$ für alle $n \in \mathbb{N}$. Dann ist die Folge T_n von Schätzfunktionen für θ konsistent im quadratischen Mittel.	■	☐
4.	Setzt man den aus einer Realisation $x_1, \dots, x_n$ einer einfachen Stichprobe nach der Maximum-Likelihood-Methode erhaltenen Parameterschätzwert in die zugehörige Likelihoodfunktion ein, so ist es möglich, dass man dabei den Wert 0 erhält.	☐	■
5.	Beim t-Test für den Erwartungswert einer normalverteilten Zufallsvariablen mit unbekannter Varianz gibt der p-Wert die minimale Abweichung des tatsächlichen Erwartungswerts vom hypothetischen Wert μ_0 an.	☐	■
6.	Die Gütefunktion eines Gauß-Tests gibt zu jedem möglichen Erwartungswert μ an, mit welcher Wahrscheinlichkeit der Test die Nullhypothese ablehnt, falls μ der zur tatsächlichen (Normal-)Verteilung von Y gehörende Erwartungswert ist.	■	☐
7.	Mit einem Chi-Quadrat-Anpassungstest soll anhand einer einfachen Stichprobe vom Umfang $n = 150$ überprüft werden, ob die Grundgesamtheit normalverteilt mit Erwartungswert $\mu = 10$ und Varianz $\sigma^2 = 2^2$ ist. Bei Verwendung einer geeigneten Klassierung aus 6 Klassen ist damit zur Konstruktion des Ablehnungsbereiches die χ^2-Verteilung mit 5 Freiheitsgraden zu verwenden.	■	☐
8.	Im einfachen linearen Regressionsmodell $$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$ sind (bei festem x_0) Prognoseintervalle für y_0 gegeben x_0 zur Vertrauenswahrscheinlichkeit $1 - \alpha$ stets breiter als die analogen Prognoseintervalle für $\mathrm{E}(y_0)$ gegeben x_0.	■	☐

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. Auf Grundlage einer einfachen Stichprobe $X_1, \dots, X_n$ vom Umfang n zu einer Zufallsvariablen Y, von der man lediglich weiß, dass sie normalverteilt ist, soll mit einem Signifikanztest überprüft werden, ob $\mathrm{Var}(Y) \neq 3^2$ gilt. Zur Untersuchung dieser Fragestellung ist geeignet:
   (a) Der χ^2-Test für die Varianz bei unbekanntem Erwartungswert ■
   (b) Der F-Test zum Varianzvergleich ☐
   (c) Die einfache Varianzanalyse ☐
   (d) Keines der in der Vorlesung besprochenen Verfahren ☐

2. Bei der Durchführung einer einfachen Varianzanalyse mit $k = 5$ Faktorstufen und einem Gesamtstichprobenumfang von $n = 255$ erhält man die realisierte Teststatistik $F = 2.52$. 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 darstellt.



3. Als p-Wert zur realisierten Teststatistik eines linksseitigen Gauß-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz (mit $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$) erhält man $p = 0.0238$. Dann gilt für die p-Werte des rechtsseitigen Tests (mit $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$) bzw. des zweiseitigen Tests (mit $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$):

   (a) Der p-Wert des rechtsseitigen Tests beträgt 0.0476, der p-Wert des zweiseitigen Tests 0.0119. ☐

   (b) Der p-Wert des rechtsseitigen Tests beträgt 0.9762, der p-Wert des zweiseitigen Tests 0.0476. ■

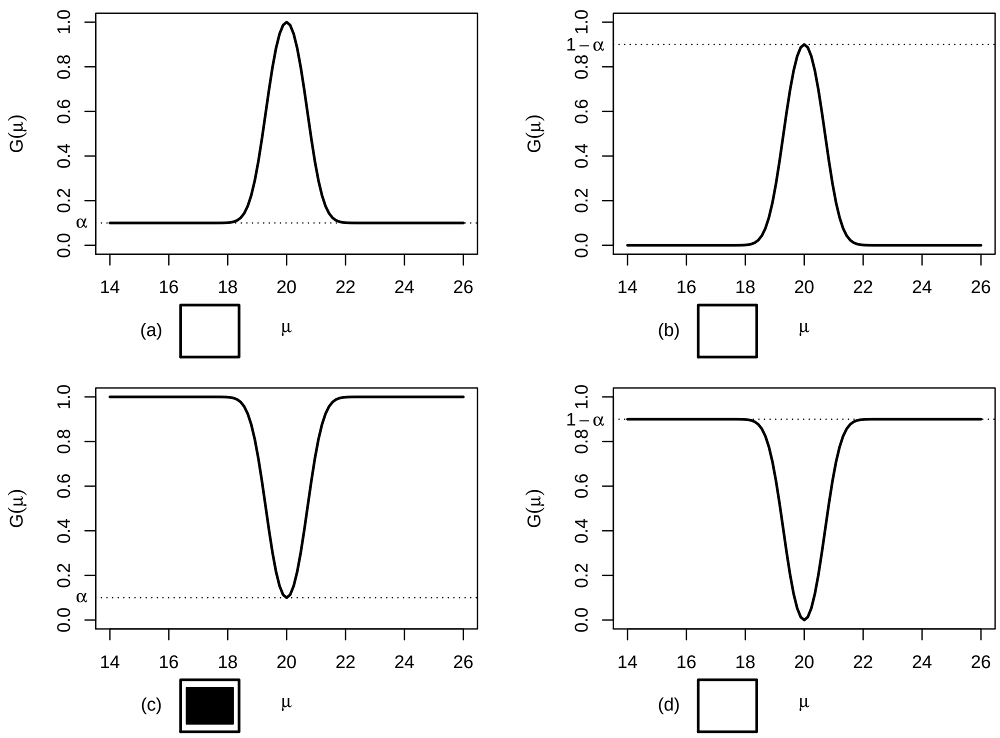
   (c) Der p-Wert des rechtsseitigen Tests beträgt 0.0119, der p-Wert des zweiseitigen Tests 0.0476. ☐

   (d) Der p-Wert des rechtsseitigen Tests beträgt 0.9762, der p-Wert des zweiseitigen Tests 0.0119. ☐

4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{49}$ vom Umfang $n = 49$ zu einer $N(\mu, 3^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu = 20 \qquad \text{gegen} \qquad H_1 : \mu \neq 20$$

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



Aufgabe 3 (4 + 2 = 6 Punkte)

In Abhängigkeit eines unbekannten Parameters $p > 0$ seien der Erwartungswert und die Varianz einer Zufallsvariablen Y gegeben durch $\mathrm{E}(Y) = \frac{2}{p}$ sowie $\mathrm{Var}(Y) = \frac{2}{p^2}$. Für $n \in \mathbb{N}$ sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y.

(a) Zeigen Sie, dass die Schätzfunktionen

$$T_n(X_1, \ldots, X_n) := \frac{1}{4n} \sum_{i=1}^{n} X_i^2$$

nicht erwartungstreu für **die Varianz von** Y sind.

(b) Geben Sie für **die Varianz von** Y erwartungstreue Schätzfunktionen $\widetilde{T}_n(X_1, \ldots, X_n)$ an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) Beweis durch Berechnung des Erwartungswerts von $T_n(X_1, \ldots, X_n)$.

(b) $\widetilde{T}_n(X_1, \ldots, X_n) = \frac{1}{3n} \sum_{i=1}^{n} X_i^2$.

Aufgabe 4 (6 + 3 + 2 = 11 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $a > 1$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|a) = \begin{cases} \dfrac{a}{y^{a+1}} & \text{für } y \geq 1 \\ 0 & \text{sonst} \end{cases}$$

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

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

(b) Zeigen Sie, dass $E(Y) = \dfrac{a}{a-1}$ gilt.

(c) Bestimmen Sie den Schätzer $\widehat{a}_{MM}$ 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.*

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{a}_{ML} = \dfrac{n}{\sum_{i=1}^{n} \ln x_i}$

(b) Nachweis durch Berechnung des Erwartungswerts

(c) $\widehat{a}_{MM} = \dfrac{\overline{x}}{\overline{x} - 1}$

Aufgabe 5 (7 + 2 + 4 = 13 Punkte)

Eine Maschine produziert Schrauben, deren Länge erfahrungsgemäß normalverteilt mit einer Standardabweichung von $0.1[cm]$ um den tatsächlichen Erwartungswert schwankt. Die laufende Qualitätskontrolle soll eine Überschreitung dieses Erwartungswerts gegenüber der mittleren Soll-Länge $6[cm]$ mit Hilfe eines geeigneten statistischen Testverfahrens auf Basis der Realisation einer einfachen Stichprobe $x_1, \ldots, x_{16}$ aufdecken. Dabei darf eine derartige Überschreitung nur mit einer Wahrscheinlichkeit von höchstens 5% fälschlicherweise signalisiert werden. Aus dem realisierten Stichprobenergebnis erhält man den Stichprobenmittelwert

$$\overline{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 6.0529[cm] \; .$$

(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** eine Überschreitung signalisieren, wenn der tatsächliche Erwartungswert der Länge der Schrauben $6.05[cm]$ beträgt?

Ergebnisse (ohne Begründung/Rechenweg):

(a) $N = 2.116 \in (1.645, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

Der Test zur Qualitätskontrolle wird also eine Überschreitung der mittleren Länge signalisieren.

(b) p-Wert $p = 0.017$. Der Test zur Qualitätskontrolle hätte also keine Überschreitung der mittleren Länge signalisiert.

(c) $G(6.05) = 0.8023$

Aufgabe 6 (11 + 9 = 20 Punkte)

Um zu überprüfen, ob sich die Leistungsfähigkeit von Lithium-Ionen-Akkus zweier verschiedener Hersteller unterscheidet, lässt ein Testinstitut die Ausdauer jeweils eines Akkus in 8 unterschiedlichen Actionkameras untersuchen. Es wurden dabei die folgenden Aufzeichnungsdauern (in Minuten) bis zur automatischen Abschaltung der Kameras festgestellt:

Kamera i	1	2	3	4	5	6	7	8
Hersteller A x_i^A	87	54	80	63	64	57	65	77
Hersteller B x_i^B	97	61	96	67	66	77	70	68

(a) Überprüfen Sie unter der Annahme, dass die so gemessenen Aufzeichnungsdauern (in Minuten) aus einer einfachen Stichprobe zur zweidimensional normalverteilten Grundgesamtheit (Y^A, Y^B) der Aufzeichnungsdauern mit Akkuhersteller A (Y^A) bzw. Akkuhersteller B (Y^B) stammen, zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass die Verwendung von Akkuhersteller A im Vergleich zu Akkuhersteller B durchschnittlich eine kürzere Aufzeichnungsdauer ermöglicht. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

(b) Nehmen Sie nun an, dass durch eine grobe Unachtsamkeit bei der Datenerhebung die Zuordnung der einzelnen Aufzeichnungsdauern zu den jeweiligen Kameramodellen verloren gegangen ist. Um die Situation zu retten, nehme man weiter an, dass mit $X_1^A, \ldots, X_8^A$ und $X_1^B, \ldots, X_8^B$ nun zwei unabhängige einfache Stichproben zu den beiden (normalverteilten) Zufallsvariablen Y^A und Y^B vorliegen. Testen Sie unter der Annahme der Varianzgleichheit von Y^A und Y^B zum Signifikanzniveau $\alpha = 0.05$ auf dieser Basis die Hypothese, dass die Verwendung von Akkuhersteller A im Vergleich zu Akkuhersteller B eine kürzere Aufzeichnungsdauer ermöglicht. Verwenden Sie hierzu die Stichprobenmittelwerte $\overline{x^A} = 68.38$ bzw. $\overline{x^B} = 75.25$ sowie die Stichprobenvarianzen $s^2_{Y^A} = 135.98$ bzw. $s^2_{Y^B} = 191.93$. Fassen Sie das Ergebnis des Tests auch in einem Antwortsatz zusammen.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $t = -2.188 \in (-\infty, -1.895) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

Der Test kann also die Vermutung, dass die Verwendung von Akkuhersteller A im Vergleich zu Akkuhersteller B eine kürzere Aufzeichnungsdauer ermöglicht, bestätigen.

(b) $t = -1.074 \notin (-\infty, -1.761) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!

Der Test kann also die Vermutung, dass die Verwendung von Akkuhersteller A im Vergleich zu Akkuhersteller B eine kürzere Aufzeichnungsdauer ermöglicht, nicht bestätigen.

Aufgabe 7 (15 Punkte)

Um zu überprüfen, ob es einen Zusammenhang zwischen der Anzahl bearbeiteter Zusatzübungsblätter und dem Abschneiden in der Klausur (bestanden/nicht bestanden) gibt, hat der Dozent einer Statistik-Veranstaltung aus den Korrekturergebnissen der zugehörigen Klausuren aller 217 Teilnehmer die folgende Tabelle zusammengestellt:

	0 Blätter bearbeitet	1 Blatt bearbeitet	2 Blätter bearbeitet
bestanden	71	43	69
nicht bestanden	25	8	1

Überprüfen Sie anhand dieses Datenmaterials zum Signifikanzniveau $\alpha = 0.05$, ob die Anzahl bearbeiteter Zusatzübungsblätter und das Klausurergebnis stochastisch unabhängig sind.

Hinweis: Verwenden Sie den folgenden Tabellenausschnitt mit p-Quantilen von $\chi^2(n)$-Verteilungen:

$n\backslash p$	0.01	0.025	0.05	0.50	0.90	0.95	0.975	0.99
1	0.000	0.001	0.004	0.455	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	1.386	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	2.366	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	3.357	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	4.351	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	5.348	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	6.346	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	7.344	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	8.343	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	9.342	15.987	18.307	20.483	23.209

Ergebnisse (ohne Begründung/Rechenweg):

$\chi^2 = 18.56 \in (5.991, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
Der Test kommt also zum Ergebnis, dass die Anzahl bearbeiteter Zusatzübungsblätter und das Klausurergebnis nicht stochastisch unabhängig sind.

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

Zur Erklärung der Höhe der deutschen Warenausfuhr y_i (in Milliarden Euro) durch die Höhe der deutschen Wareneinfuhr x_i (in Milliarden Euro) 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 vorliegenden Daten zu den Jahren 2010–2016 wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
      1       2       3       4       5       6       7
 14.980 -44.635  -8.270   1.545   5.693  12.968  17.719

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -339.6649   167.7718  -2.025 0.098793 .
x              1.6017     0.1861   8.609 0.000349 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.67 on 5 degrees of freedom
Multiple R-squared:  0.9368,	Adjusted R-squared:  0.9242
F-statistic: 74.11 on 1 and 5 DF,  p-value: 0.000349
```

(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 Höhe der deutschen Warenausfuhr wird durch das lineare Modell erklärt?

(d) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.05$, ob β_1 signifikant von Null verschieden ist.

(e) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.01$, ob β_2 signifikant positiv ist.

(f) Welche Warenausfuhrhöhe (in Milliarden Euro) prognostiziert das Modell für ein Jahr mit einer Wareneinfuhrhöhe von 950 (in Milliarden Euro)?

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_1 = -339.6649$, $\widehat{\beta}_2 = 1.6017$

(b) $\widehat{\sigma^2} = 560.2689$

(c) 0.9368

(d) β_1 ist nicht signifikant von Null verschieden.

(e) β_2 ist signifikant positiv.

(f) 1181.9501

Aufgabe 9 (6 + 2 + 3 + 5 + 5 = 21 Punkte)

Zur Schätzung eines einfachen linearen Regressionsmodells

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

aus einer Stichprobe vom Umfang $n = 25$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{25} y_i = 174.003; \quad \sum_{i=1}^{25} y_i^2 = 1664.405; \quad \sum_{i=1}^{25} x_i = 112.315;$$

$$\sum_{i=1}^{25} x_i^2 = 548.684; \quad \sum_{i=1}^{25} x_i \cdot y_i = 662.334$$

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

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

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

(d) Überprüfen Sie zum Signifikanzniveau $\alpha = 0.05$, ob β_2 signifikant negativ ist. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

(e) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für $\mathrm{E}(y_0)$ gegeben $x_0 = 6$ an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_1 = 19.1234$, $\widehat{\beta}_2 = -2.7074$

(b) $\widehat{\sigma}^2 = 5.6565$

(c) $\widehat{\sigma^2}_{\widehat{\beta}_1} = 2.8152$, $\widehat{\sigma^2}_{\widehat{\beta}_2} = 0.12827$

(d) $t = -7.56 \in (-\infty, -1.714) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

β_2 ist also signifikant negativ.

(e) $[1.39, 4.368]$