Wirtschaftswissenschaftliches Prüfungsamt

Bachelor-Prüfung
Schließende Statistik
Wintersemester 2012/13

Aufgabenstellung und Ergebnisse

Dr. Martin Becker

Hinweise für die Klausurteilnehmer

- Kleben Sie bitte **sofort** Ihr Namensschild auf obige Markierung!
- 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 + 10 + 16 + 10 + 14 + 16 + 6 + 20) 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.

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 −1 Punkt, nicht bearbeitete Aussagen 0 Punkte (Aussagen mit zwei Kreuzen zählen als nicht bearbeitet!).
Die Aufgabe wird insgesamt mit mindestens 0 Punkten bewertet!

	Aussage	wahr	falsch
1.	Sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu einer normalverteilten Zufallsvariablen Y. Dann sind $X_1, \ldots, X_n$ stets unkorreliert.	■	□
2.	Sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu einer $N(\mu, \sigma^2)$-verteilten Zufallsvariablen Y, sei $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Dann gilt für $\mu_0 \in \mathbb{R}$ stets $\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \sim N(\frac{\mu-\mu_0}{\sigma}\sqrt{n}, 1)$.	■	□
3.	Wird bei einem Alternativtest die Nullhypothese abgelehnt, obwohl sie gültig ist, handelt es sich um einen Fehler 2. Art.	□	■
4.	Bei einem statistischen Alternativtest zum Signifikanzniveau α ist die minimale Wahrscheinlichkeit für einen Fehler 2. Art stets größer oder gleich α.	□	■
5.	Die Gütefunktion $G(\mu)$ nimmt bei zweiseitigen Gauß-Tests für den Erwartungswert einer normalverteilten Zufallsvariablen bei bekannter Varianz zum Signifikanzniveau α mit $H_0 : \mu = \mu_0$ unabhängig vom Stichprobenumfang n an der Stelle μ_0 stets den Wert α an.	■	□
6.	Sind die Voraussetzungen für die exakte Anwendung des 2-Stichproben-t-Tests für den Vergleich zweier Erwartungswerte erfüllt, so unterscheiden sich die Verteilungen von Y^A und Y^B höchstens in der Lage.	■	□
7.	Sind die Voraussetzungen zur Anwendung der (einfaktoriellen) Varianzanalyse erfüllt und liegen als Stichprobenrealisation jeweils 25 Beobachtungen zu 4 Faktorstufen vor, so ist die Teststatistik unter der Nullhypothese $F(3, 97)$-verteilt.	□	■
8.	Im einfachen linearen Regressionsmodell $y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, haben die aus der Vorlesung bekannten Schätzfunktionen $\widehat{\beta}_1$ und $\widehat{\beta}_2$ unter allen (in y_i) linearen und (für β_1 bzw. β_2) erwartungstreuen Schätzfunktionen die kleinste Varianz.	■	□

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 −1 Punkt, nicht bearbeitete Aufgabenteile 0 Punkte (Aufgabenteile mit mehr als einem Kreuz zählen als nicht bearbeitet!).
Die Aufgabe wird insgesamt mit mindestens 0 Punkten bewertet!

1. Eine Familie $\widehat{\theta}_n, n \in \mathbb{N}$, von Schätzfunktionen für einen Parameter $\theta \in \Theta \subseteq \mathbb{R}$ ist **genau** dann konsistent für θ, wenn (für alle $\theta \in \Theta$)
   - (a) $\mathrm{E}(\widehat{\theta}_n) = \theta$ für jedes $n \in \mathbb{N}$ und außerdem $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) > 0$ gilt. ☐
   - (b) $\mathrm{E}(\widehat{\theta}_n) = \theta$ für jedes $n \in \mathbb{N}$ und außerdem $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) = 0$ gilt. ☐
   - (c) $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n) = \theta$ und außerdem $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) > 0$ gilt. ☐
   - (d) $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n) = \theta$ und außerdem $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) = 0$ gilt. ■

2. Mit Hilfe einer einfachen Stichprobe $X_1, \ldots, X_n$ vom Umfang n zu einer $N(\mu, \sigma^2)$-verteilten Zufallsvariablen mit unbekanntem Erwartungswert μ soll mit dem Chi-Quadrat-Test für die Varianz zwischen den Hypothesen $H_0 : \sigma^2 \geq 4$ und $H_1 : \sigma^2 < 4$ entschieden werden. Dann gilt für den kritischen Bereich K zum Signifikanzniveau α:
   - (a) $K = [-\infty, -\chi^2_{n-1,1-\alpha})$ ☐
   - (b) $K = [-\infty, -\chi^2_{n,1-\alpha})$ ☐
   - (c) $K = [0, \chi^2_{n-1,\alpha})$ ■
   - (d) $K = [0, \chi^2_{n,\alpha})$ ☐

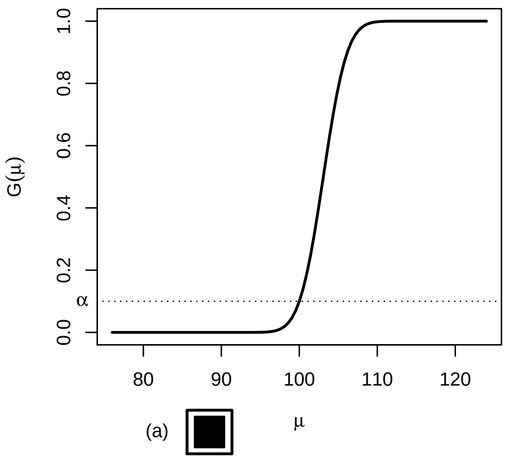
3. Bei der Durchführung eines t-Tests für den Erwartungswert einer normalverteilten Zufallsvariablen bei unbekannter Varianz auf Grundlage einer einfachen Stichprobe vom Umfang n zum Signifikanzniveau α lehnt der linksseitige Test H_0 ab, während der zweiseitige Test H_0 nicht verwerfen kann. Damit weiß man über die Realisation t der Teststatistik:
   - (a) $t \in (-\infty, -t_{n-1,1-\frac{\alpha}{2}})$ ☐
   - (b) $t \in [-t_{n-1,1-\frac{\alpha}{2}}, -t_{n-1,1-\alpha})$ ■
   - (c) $t \in [-t_{n-1,1-\alpha}, t_{n-1,1-\alpha}]$ ☐
   - (d) $t \in (t_{n-1,1-\alpha}, \infty)$ ☐

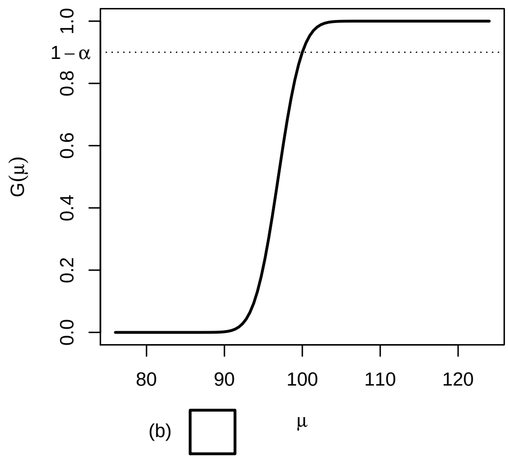
4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{25}$ vom Umfang $n = 25$ zu einer $N(\mu, 12^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

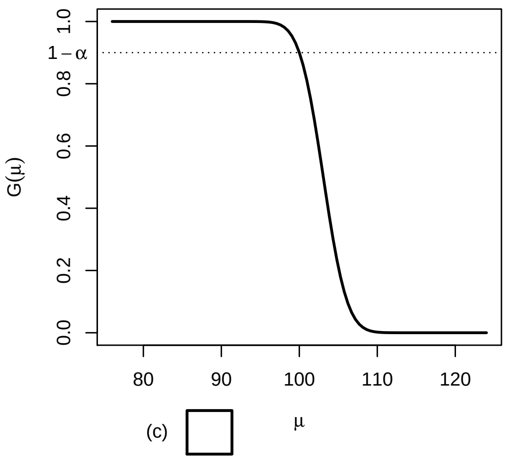
$$H_0 : \mu \leq 100 \qquad \text{gegen} \qquad H_1 : \mu > 100$$

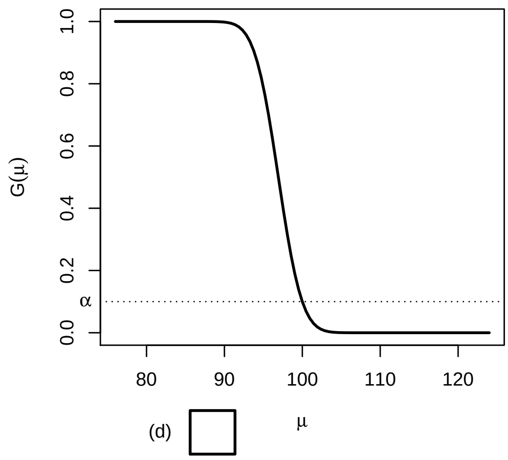
bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.

Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.









Aufgabe 3 (6 + 2 + 2 = 10 Punkte)

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

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

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 $\mathrm{E}(Y) = \dfrac{a}{a-1}$ gilt.

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

Hinweise:

- *Beachten Sie, dass Sie Teil (c) auch ohne die Bearbeitung von Teil (a) und Teil (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 4 (7 + 2 + 4 + 3 = 16 Punkte)

Bei der Abfüllung von Tiefengrund weiß der Hersteller aus langjähriger Erfahrung, dass die verwendete Maschine eine Standardabweichung von $0.05[l]$ für die abgefüllte Menge hat. Nach einer routinemäßigen Überprüfung hat der Hersteller den Verdacht, dass die Abfüllanlage im Mittel weniger als die auf dem Produkt ausgezeichneten $5[l]$ in die Kanister einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 36 Behälter entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{36}$ als Realisation einer einfachen Stichprobe vom Umfang 36 zur annahmegemäß $N(\mu, 0.05^2[l^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{36} \sum_{i=1}^{36} x_i = 4.983[l] \ .$$

(a) Testen Sie zum Signifikanzniveau $\alpha = 0.025$ **(!)**, ob der Verdacht der Herstellerfirma bestätigt werden kann. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Berechnen Sie den p-Wert zum Test aus Teil (a).

(c) Wie groß ist die Wahrscheinlichkeit des Fehlers 2. Art, falls $\mu = 4.97[l]$ beträgt?

(d) Geben Sie auf Basis der Stichprobenrealisation ein zweiseitiges Konfidenzintervall für die mittlere Abfüllmenge zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.95$ an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $N = -2.04 \in (-\infty, -1.96) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

Der Test bestätigt also den Verdacht der Herstellerfirma, dass die von der Maschine abgefüllte Menge im Mittel zu niedrig ist.

(b) $p = 0.0207$

(c) $\beta(4.97) = 0.0505$

(d) Realisiertes symm. Konfidenzintervall zum Konf.-niveau $1 - \alpha = 0.95$: $[4.967, 4.999]$

Aufgabe 5 (10 Punkte)

Zwei unterschiedlichen Gruppen mit jeweils 61 Schmerzpatienten wird jeweils ein spezielles Schmerzmittel verabreicht. Nach einer festgelegten Zeit werden dann alle Patienten gefragt, ob durch das verabreichte Schmerzmittel eine spürbare Linderung der Schmerzen eingetreten ist. In der Gruppe der Patienten, denen Schmerzmittel A verabreicht wurde, beantworten 36 Personen diese Frage positiv, in der zu Schmerzmittel B gehörigen Gruppe 29 Personen. Überprüfen Sie unter der Annahme, dass es sich bei dem Stichprobenergebnis um die Realisation zweier unabhängiger einfacher Stichproben handelt, zum Signifikanzniveau $\alpha = 0.05$, ob sich die Wirksamkeit der beiden Schmerzmittel unterscheidet (bezogen auf die Erfolgswahrscheinlichkeit für eine Linderung der Schmerzen). Formulieren Sie das Ergebnis auch in Form eines Antwortsatzes.

Ergebnisse (ohne Begründung/Rechenweg):

$t = 1.2688 \notin (-\infty, -1.98) \cup (1.98, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
Der Test findet also keine Anzeichen dafür, dass sich die Wirksamkeit der beiden Schmerzmittel unterscheidet.

Aufgabe 6 (14 Punkte)

Mit einem Chi-Quadrat-Anpassungstest soll zum Signifikanzniveau $\alpha = 0.05$ getestet werden, ob die von einem Zufallszahlengenerator erzeugten Zufallszahlen (wie gewünscht) Exp(1)-verteilt sind. Dazu wurden $n = 100$ unabhängige Zufallszahlen generiert und die Verteilung auf einer vorgegebenen Intervalleinteilung wie folgt festgestellt:

i	1	2	3	4
K_i	$(-\infty, 0.5]$	$(0.5, 1]$	$(1, 2]$	$(2, \infty)$
n_i	32	31	28	9

Führen Sie den beschriebenen Test durch. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

Hinweise:

- *Die Verteilungsfunktion einer* Exp(λ)*-verteilten Zufallsvariablen* X *ist gegeben durch:*

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } x \geq 0 \end{cases}$$

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

$n \backslash p$	0.01	0.025	0.05	0.5	0.9	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
96	66.730	70.783	74.401	95.334	114.131	119.871	125.000	131.141
97	67.562	71.642	75.282	96.334	115.223	120.990	126.141	132.309
98	68.396	72.501	76.164	97.334	116.315	122.108	127.282	133.476
99	69.230	73.361	77.046	98.334	117.407	123.225	128.422	134.642
100	70.065	74.222	77.929	99.334	118.498	124.342	129.561	135.807

Ergebnisse (ohne Begründung/Rechenweg):

$\chi^2 = 5.9921 \notin (7.815, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
Der Test findet also keine Anzeichen dafür, dass die vom Zufallszahlengenerator erzeugten Zahlen nicht Exp(1)-verteilt sind.

Aufgabe 7 (16 Punkte)

Um zu überprüfen, ob es einen Zusammenhang zwischen dem Geschlecht des Studierenden (männlich/weiblich) und dem erreichten Klausurergebnis (Zugehörigkeit zu einer Notengruppe) gibt, hat der Dozent einer Statistik-Veranstaltung aus den Korrekturergebnissen der zugehörigen Klausuren aller 195 Teilnehmer die folgende Tabelle zusammengestellt:

	Note 1.0 – 2.3	Note 2.7 – 4.0	Note 5.0
männlich	27	74	30
weiblich	15	39	10

Überprüfen Sie anhand dieses Datenmaterials zum Signifikanzniveau $\alpha = 0.05$, ob Geschlecht und Klausurergebnis stochastisch abhängig sind.

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

$n\backslash p$	0.01	0.025	0.05	0.5	0.9	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):

$\chi^2 = 1.416 \notin (5.991, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
Der Test kommt also zum Ergebnis, dass Geschlecht und Klausurergebnis stochastisch unabhängig sind.

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

Zur Erklärung der (mittleren) Geburtenrate y_i (in Kinder pro Frau) durch das (mittlere) jährliche Pro-Kopf Bruttoinlandsprodukt x_i (in 1 000 USD) 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 Vereinten Nationen (für Länder mit einem Pro-Kopf BIP von mehr als 10 000 USD) des Jahres 1998 wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
    Min      1Q  Median      3Q     Max
-1.1399 -0.4014 -0.1413  0.2616  2.7818

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.90511    0.31848   9.122 6.82e-11 ***
x           -0.03864    0.01331  -2.903  0.00627 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.723 on 36 degrees of freedom
Multiple R-squared: 0.1897,       Adjusted R-squared: 0.1672
F-statistic: 8.428 on 1 and 36 DF,  p-value: 0.006274
```

(a) Wie viele Länder gingen in die Schätzung ein?

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

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

(d) Welcher Anteil der Gesamtvarianz der Geburtenrate wird durch das lineare Modell erklärt?

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

(f) Welche (mittlere) Geburtenrate prognostiziert das Modell für ein Land mit einem (mittleren) jährlichen Pro-Kopf BIP von 20 (in 1 000 USD)?

Ergebnisse (ohne Begründung/Rechenweg):

(a) $n = 38$

(b) $\widehat{\beta}_1 = 2.90511$, $\widehat{\beta}_2 = -0.03864$

(c) $\widehat{\sigma}^2 = 0.5227$

(d) 0.1897

(e) β_2 signifikant negativ.

(f) $\widehat{y}_0 = 2.132$

Aufgabe 9 (6 + 3 + 3 + 5 + 3 = 20 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 $n = 25$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{25} y_i = 280.923; \quad \sum_{i=1}^{25} y_i^2 = 3358.773; \quad \sum_{i=1}^{25} x_i = 148.569;$$

$$\sum_{i=1}^{25} x_i^2 = 954.92; \quad \sum_{i=1}^{25} x_i \cdot y_i = 1571.658; \quad \sum_{i=1}^{25} \hat{y}_i^2 = 3289.561.$$

(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 $\alpha = 0.01$, ob β_2 signifikant negativ ist. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

(e) Geben Sie ein symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_1 an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_1 = 19.319$, $\widehat{\beta}_2 = -1.36$

(b) $\widehat{\sigma}^2 = 2.996$

(c) $\widehat{\sigma^2}_{\widehat{\beta}_1} = 1.591$, $\widehat{\sigma^2}_{\widehat{\beta}_2} = 0.04164$

(d) $t = -6.663 \in (-\infty, -2.5) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

β_1 ist also signifikant negativ.

(e) $[16.71, 21.928]$