Aufgabenstellung und Ergebnisse zur

Bachelor-Prüfung
Schliessende Statistik
Sommersemester 2020

PD Dr. Martin Becker

Hinweise für die Klausurteilnehmer

- Die Klausur besteht aus insgesamt 10 Aufgaben. Prüfen Sie die Vollständigkeit Ihres Exemplares nach; spätere Reklamationen können nicht berücksichtigt werden.
- Es sind insgesamt 120 Punkte (= 16 + 12 + 8 + 12 + 16 + 11 + 10 + 14 + 5 + 16) 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 1–12 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)	Σ
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
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.	Ist $X_1, \dots, X_n$ eine einfache Stichprobe vom Umfang n zu einer normalverteilten Zufallsvariablen Y, dann sind Schwankungsintervalle für $\overline{X}$ umso breiter, je größer der Stichprobenumfang n ist.	□	■
2.	Sind $\widehat{\theta}$ und $\widetilde{\theta}$ zwei für $\theta \in \Theta$ erwartungstreue Schätzfunktionen und ist $\widehat{\theta}$ wirksamer als $\widetilde{\theta}$, so ist die Varianz von $\widehat{\theta}$ für kein einziges $\theta \in \Theta$ größer als die von $\widetilde{\theta}$.	■	□
3.	Die Nullhypothese $H_0 : \mu = 42$ wird beim Gauß-Test auf den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz zum Signifikanzniveau α genau dann abgelehnt, wenn 42 nicht im entsprechenden (symmetrischen) Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha$ für μ enthalten ist.	■	□
4.	Erhöht man bei einem linksseitigen Gauß-Test auf den Mittelwert einer normalverteilten Zufallsvariablen bei bekannter Varianz den Stichprobenumfang n, so verringert man damit (mit Ausnahme der Situation $\mu = \mu_0$) sowohl die Fehlerwahrscheinlichkeiten 1. Art als auch die Fehlerwahrscheinlichkeiten 2. Art.	■	□
5.	Das Vergrößern des Signifikanzniveaus α führt bei sämtlichen in der Veranstaltung besprochenen Hypothesentests stets zu einer Vergrößerung des kritischen Bereichs.	■	□
6.	Kann ein Chi-Quadrat-Unabhängigkeitstest die Nullhypothese zum Signifikanzniveau $\alpha = 0.10$ nicht ablehnen, so wird die Nullhypothese auch bei einem Signifikanzniveau von $\alpha = 0.05$ nicht verworfen.	■	□
7.	Stimmen bei der Anwendung des zweiseitigen F-Tests zum Vergleich der Varianzen zweier normalverteilter Zufallsvariablen Y^A und Y^B die beiden Stichprobengrößen n_A und n_B überein, so ändert sich der kritische Bereich nicht, wenn man Y^A und Y^B vertauscht.	■	□
8.	Im einfachen linearen Regressionsmodell $$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2),$$ sind Prognoseintervalle für $\mathrm{E}(y_0)$ gegeben x_0 umso breiter, je weiter x_0 von 0 entfernt ist.	□	■

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!).

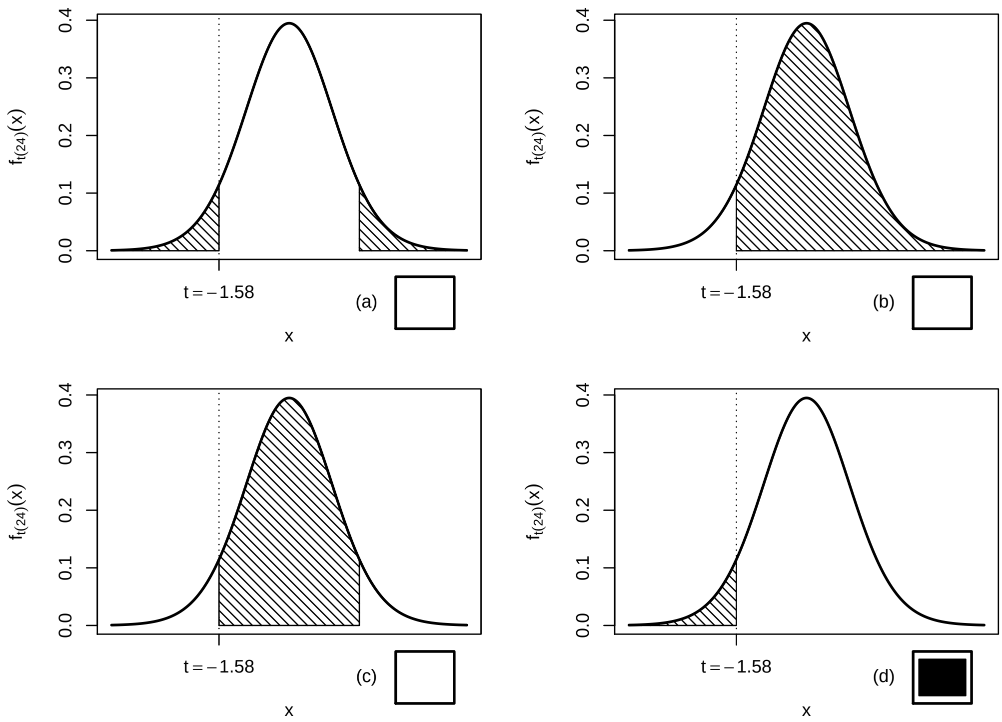
1. Bei der Durchführung einer einfachen Varianzanalyse wurden zu den 3 Faktorstufen jeweils einfache Stichproben mit den Stichprobenumfängen 20, 30 beziehungsweise 40 erhoben. Damit besitzt die Teststatistik bei Gültigkeit der Nullhypothese (und sämtlicher Voraussetzungen zur exakten Anwendungsmöglichkeit des Tests) die folgende Verteilung:

   (a) $F(3, 90)$ ☐
   (b) $F(2, 90)$ ☐
   (c) $F(3, 87)$ ☐
   (d) $F(2, 87)$ ■

2. Sei $X_1, \ldots, X_{25}$ 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 $n = 25$ soll

$$H_0 : \mu \geq \mu_0 = 10 \qquad \text{gegen} \qquad H_1 : \mu < \mu_0 = 10$$

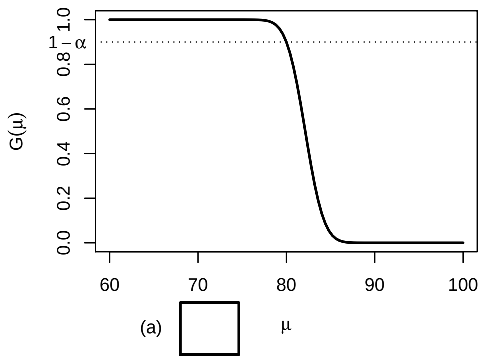
   mit einem t-Test getestet werden. Als realisierte Teststatistik erhält man $t = -1.58$. Markieren Sie die Abbildung, welche den p-Wert in der beschriebenen Situation korrekt als Inhalt der schraffierten Fläche unter der Dichtefunktion der Verteilung der Teststatistik unter H_0 (für $\mu = \mu_0$) darstellt.

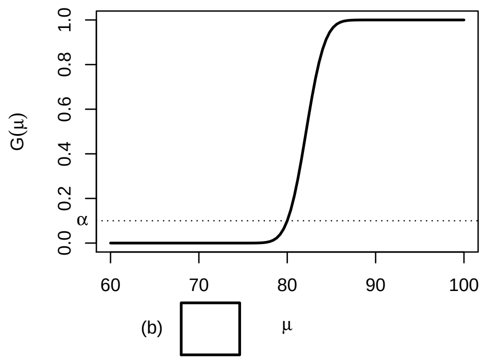


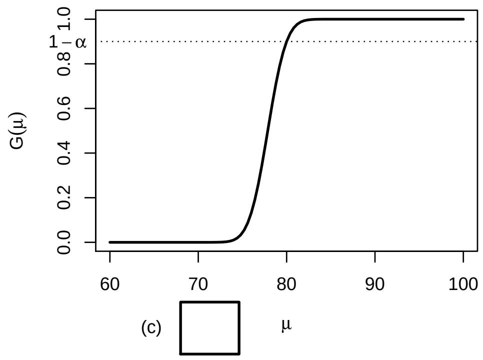
3. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{36}$ vom Umfang $n = 36$ zu einer $N(\mu, 10^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

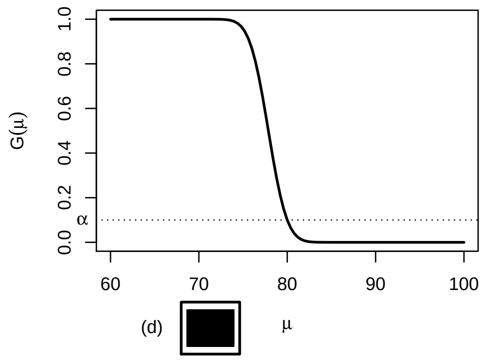
$$H_0 : \mu \geq 80 \qquad \text{gegen} \qquad H_1 : \mu < 80$$

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









4. Bei der Durchführung eines zweiseitigen Gauß-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz (mit $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$) zum Signifikanzniveau $\alpha = 0.05$ wird H_0 abgelehnt. Dann gilt für das Ergebnis der einseitigen Tests (mit $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$ bzw. $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$) zum Signifikanzniveau $\alpha = 0.05$ (auf Grundlage derselben Stichprobenrealisation):

   (a) Bei keinem der beiden einseitigen Tests wird H_0 abgelehnt. ☐

   (b) Bei genau einem der beiden einseitigen Tests wird H_0 abgelehnt. Bei welchem dies der Fall ist, hängt vom Vorzeichen der Teststatistik ab. ■

   (c) Bei beiden einseitigen Tests wird H_0 abgelehnt. ☐

   (d) Auf Grundlage der vorhandenen Informationen ist noch unklar, ob H_0 bei keinem oder genau einem einseitigen Test abgelehnt wird. ☐

Aufgabe 3 (3 + 3 + 2 = 8 Punkte)

In Abhängigkeit eines unbekannten Parameters c mit $0 \leq c \leq 10$ seien der Erwartungswert und die Varianz von Zufallsvariablen Y mit der zugehörigen Verteilung aus einer parametrischen Verteilungsfamilie gegeben durch

$$\mathrm{E}(Y) = \frac{c+10}{3} \qquad \text{sowie} \qquad \mathrm{Var}(Y) = \frac{c^2 - 10c + 100}{18} \ .$$

Für $n \in \mathbb{N}$ seien $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y und T_n die wie folgt definierte Schätzfunktion für c:

$$T_n(X_1, \ldots, X_n) := \left(\frac{3}{n} \sum_{i=1}^{n} X_i \right) - 10$$

(a) Berechnen Sie den Bias der Schätzfunktionen T_n für c.

(b) Berechnen Sie die Varianz der Schätzfunktionen T_n.

(c) Ist die Folge von Schätzfunktionen T_n, $n \in \mathbb{N}$, konsistent im quadratischen Mittel für c? *(Begründung erforderlich!)*

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\mathsf{Bias}(T_n(X_1, \ldots, X_n)) = 0$.

(b) $\mathrm{Var}(T_n(X_1, \ldots, X_n)) = \frac{c^2-10c+100}{2n}$.

(c) Ja.

<u>Aufgabe 4</u> (6 + 4 + 2 = 12 Punkte)

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

$$f_Y(y|b) = \begin{cases} \dfrac{b \cdot 2^b}{y^{b+1}} & \text{für } y \geq 2 \\ 0 & \text{sonst} \end{cases}$$

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 $E(Y) = \dfrac{2b}{b-1}$ 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.*

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{b}_{ML} = \dfrac{n}{\sum_{i=1}^{n}(\ln x_i - \ln 2)}$

(b) Nachweis durch Berechnung des Erwartungswerts

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

Aufgabe 5 $(3 + 7 + 2 + 4 = 16$ Punkte)

Bei der Abfüllung von Desinfektionsmittel weiß der Hersteller aus Erfahrung, dass die verwendete Maschine eine Standardabweichung von $5[ml]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Abfüllanlage im Mittel weniger als die auf dem Produkt ausgezeichneten $1000[ml]$ in die Euroflaschen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 25 Euroflaschen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{25}$ als Realisation einer einfachen Stichprobe vom Umfang 25 zur annahmegemäß $N(\mu, 5^2[ml^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 998.261[ml] .$$

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

(b) Testen Sie zum Signifikanzniveau $\alpha = 0.05$, ob der Verdacht der Herstellerfirma bestätigt werden kann. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

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

(d) Wie groß ist die Wahrscheinlichkeit, bei der Ziehung einer Stichprobe und der anschließenden Durchführung des Tests aus Teil (b) eine Testentscheidung zu Gunsten der Nullhypothese zu erhalten, falls die tatsächliche mittlere Abfüllmenge $997[ml]$ beträgt?

Ergebnisse (ohne Begründung/Rechenweg):

(a) Realisiertes symm. Konfidenzintervall zum Konf.-niveau $1-\alpha = 0.95$: $[996.301, 1000.221]$

(b) $N = -1.739 \in (-\infty, -1.645) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

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

(c) p-Wert $p = 0.0409$. Entscheidung wäre zu Gunsten der Nullhypothese ausgefallen.

(d) $\beta(997) = 0.0869$

Aufgabe 6 (11 Punkte)

Um zu überprüfen, ob sich die Leistungsfähigkeit von Alkali-Mangan-Batterien zweier verschiedener Marken unterscheidet, lässt ein Testinstitut die Ausdauer jeweils eines Batteriesatzes in 8 unterschiedlichen Digitalkameramodellen untersuchen. Es wurden dabei die folgenden Aufnahmeanzahlen bis zur automatischen Abschaltung der Kameras festgestellt:

Kamera i	1	2	3	4	5	6	7	8
Marke A x_i^A	288	318	341	318	312	313	314	328
Marke B x_i^B	299	338	322	312	344	321	325	362

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

Ergebnisse (ohne Begründung/Rechenweg):

$t = -1.7954 \notin (-\infty, -1.895) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
Der Test kann also die Vermutung, dass die Verwendung von Batteriemarke A im Vergleich zu Batteriemarke B eine niedrigere Aufnahmeanzahl ermöglicht, nicht bestätigen.

Aufgabe 7 (10 Punkte)

Zur Behandlung einer neuartigen Infektionskrankheit werden in einer klinischen Studie zwei Medikamente miteinander verglichen. Hierzu werden nach Diagnose der Krankheit zwei unterschiedlichen Gruppen mit 43 (Gruppe A) bzw. 39 (Gruppe B) Patienten jeweils eines der Medikamente verabreicht. Nach einer festgelegten Zeit wird dann bei allen Patienten festgestellt, ob sich der Gesundheitszustand verbessert hat. In der Gruppe der Patienten, denen Medikament A verabreicht wurde, wurde bei 31 Personen eine Verbesserung festgestellt, in der zu Medikament B gehörigen Gruppe bei 32 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 Medikamente unterscheidet (bezogen auf die Erfolgswahrscheinlichkeit für eine Besserung des Gesundheitszustands). Formulieren Sie das Ergebnis auch in Form eines Antwortsatzes.

Ergebnisse (ohne Begründung/Rechenweg):

$t = -1.0618 \notin (-\infty, -1.99) \cup (1.99, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
Der Test findet also keine Anzeichen dafür, dass sich die Wirksamkeit der beiden Medikamente unterscheidet.

Aufgabe 8 (14 Punkte)

Mit einem Hypothesentest soll überprüft werden, ob die in Form folgender Häufigkeitsverteilung vorliegende Realisation einer einfachen Stichprobe vom Umfang $n = 100$ mit der Annahme einer Normalverteilung für die zugrundeliegende Zufallsvariable Y vereinbar ist:

i	1	2	3	4
K_i	$(-\infty, 15]$	$(15, 25]$	$(25, 35]$	$(35, \infty)$
n_i	6	43	28	23

Aus der vorliegenden Stichprobenrealisation wurden bereits (gerundet) die beiden Parameter $\widehat{\mu} = 27.1$ und $\widehat{\sigma}^2 = 9.3^2$ per ML-Methode aus den klassierten Daten geschätzt. Führen Sie auf dieser Grundlage einen geeigneten Signifikanztest zum Signifikanzniveau $\alpha = 0.05$ durch!

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

Ergebnisse (ohne Begründung/Rechenweg):

$\chi^2 = 9.6355 \in (3.841, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt! [1] 0.0019
Der Test kommt also zum Ergebnis, dass die Annahme einer Normalverteilung verworfen werden muss.

Aufgabe 9 (1 + 1 + 1 + 1 + 1 = 5 Punkte)

Zur Erklärung der Gesamtzahl der bestätigten Corona-Infektionen im Saarland zu Beginn der Pandemie ab dem 11.03.2020 y_i durch die quadrierte Anzahl der seit 10.03.2020 vergangenen Tage x_i 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 des saarländischen Gesundheitsministeriums wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
    Min      1Q  Median      3Q     Max
-84.873 -13.704   3.908  13.511  83.849

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.49217    9.79330   0.663    0.513
x            2.19595    0.02336  94.016   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.38 on 28 degrees of freedom
Multiple R-squared:  0.9968,       Adjusted R-squared:  0.9967
F-statistic:  8839 on 1 and 28 DF,  p-value: < 2.2e-16
```

(a) Wie viele Tage 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 Gesamtzahl der bestätigten Corona-Infektionen im Saarland zu Beginn der Pandemie ab dem 11.03.2020 wird durch das lineare Modell erklärt?

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

Ergebnisse (ohne Begründung/Rechenweg):

(a) $n = 30$

(b) $\widehat{\beta}_1 = 6.4922$, $\widehat{\beta}_2 = 2.196$

(c) $\widehat{\sigma^2} = 1251.7444$

(d) 0.9968

(e) β_1 ist nicht signifikant positiv.

Aufgabe 10 (6 + 2 + 3 + 5 = 16 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 $n = 20$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{20} y_i = 88.795; \quad \sum_{i=1}^{20} y_i^2 = 482.068; \quad \sum_{i=1}^{20} x_i = 58.473;$$

$$\sum_{i=1}^{20} x_i^2 = 191.99; \quad \sum_{i=1}^{20} x_i \cdot y_i = 224.478$$

(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) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für $\mathrm{E}(y_0)$ gegeben $x_0 = 2$ an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_1 = 9.3244$, $\widehat{\beta}_2 = -1.6707$

(b) $\widehat{\sigma}^2 = 1.6185$

(c) $\widehat{\sigma^2}_{\widehat{\beta}_1} = 0.73879$, $\widehat{\sigma^2}_{\widehat{\beta}_2} = 0.076961$

(d) $[5.179, 6.787]$