Klausurensammlung Schließende Statistik

PD Dr. Martin Becker

13. Januar 2022

Inhaltsverzeichnis

1 Schließende Statistik WS 2014/15

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!).

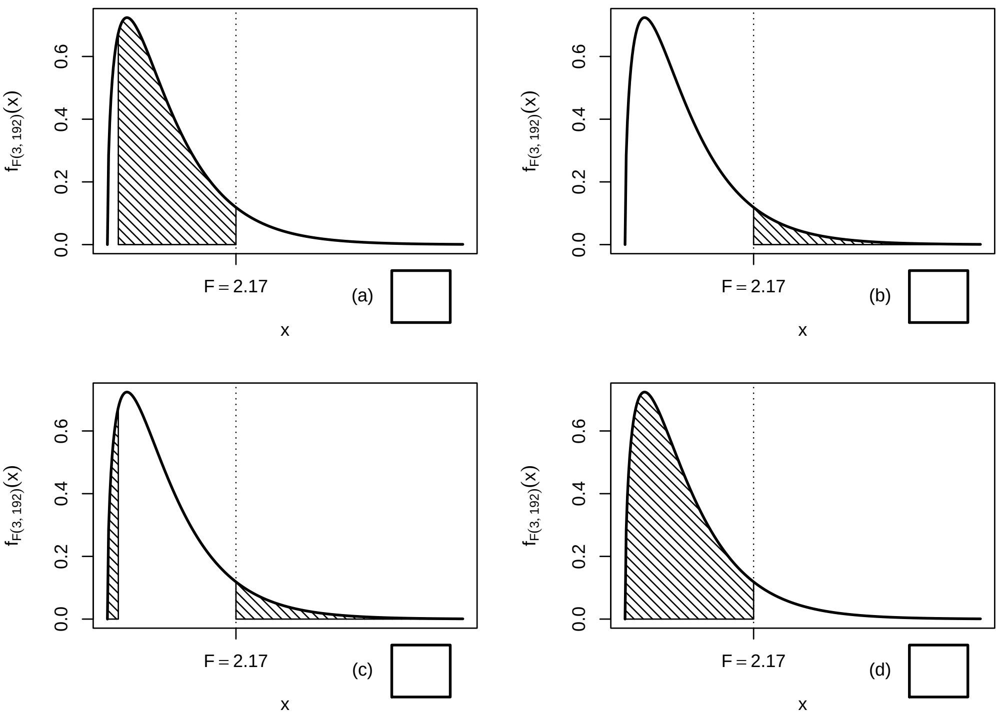
	wahr	falsch
1. Sei $X_1, \dots, X_n$ eine einfache Stichprobe vom Umfang n zu einer normalverteilten Zufallsvariablen Y. Dann besitzen $X_1, \dots, X_n$ stets übereinstimmende Erwartungswerte und übereinstimmende Varianzen.	☐	☐
2. Ist eine Familie von Schätzfunktionen $T_n, n \in \mathbb{N}$, konsistent im quadratischen Mittel für einen Parameter $\lambda \in \mathbb{R}$, so gilt stets $$\lim_{n\to\infty} \mathrm{Var}(T_n) = 0.$$	☐	☐
3. Die Breite von Konfidenzintervallen für den Erwartungswert μ einer normalverteilten Zufallsvariablen bei bekannter Varianz vergrößert sich mit wachsender Varianz.	☐	☐
4. Das Verkleinern des Signifikanzniveaus eines Hypothesentests führt stets zu einer Vergrößerung des kritischen Bereichs.	☐	☐
5. Gilt bei der Anwendung eines statistischen Tests für den p-Wert p zur Teststatistik T die Beziehung $p < \alpha$, so liegt T im kritischen Bereich des zugehörigen Tests zum Signifikanzniveau α.	☐	☐
6. Nimmt ein zweiseitiger Gauß-Test die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ an, so wird H_0 stets auch zum Signifikanzniveau $\alpha = 0.10$ angenommen.	☐	☐
7. Sind die Voraussetzungen zur Anwendung der einfachen Varianzanalyse erfüllt und liegen als Stichprobenrealisation jeweils 20 Beobachtungen zu 5 Faktorstufen vor, so ist die Teststatistik bei Gültigkeit der Nullhypothese $F(5, 95)$-verteilt.	☐	☐
8. Die Anwendung der Kleinst-Quadrate-Methode bedeutet, die Summe der quadrierten horizontalen Abstände der Beobachtungspunkte zur Regressionsgeraden zu maximieren.	☐	☐

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. Beim zweiseitigen Gauß-Test für den Erwartungswert einer normalverteilten Zufallsvariablen mit bekannter Varianz zum Signifikanzniveau $\alpha \in (0, 0.5)$
   (a) liegt die Teststatistik N bei Gültigkeit von H_0 nie im kritischen Bereich K. ☐
   (b) liegt die Teststatistik N bei Gültigkeit von H_0 mit Wahrscheinlichkeit α im kritischen Bereich K. ☐
   (c) liegt die Teststatistik N bei Gültigkeit von H_0 mit Wahrscheinlichkeit $1 - \alpha$ im kritischen Bereich K. ☐
   (d) liegt die Teststatistik N bei Gültigkeit von H_0 stets im kritischen Bereich K. ☐

2. Bei der Durchführung einer einfachen Varianzanalyse mit $k = 4$ Faktorstufen und einem Gesamtstichprobenumfang von $n = 196$ erhält man die realisierte Teststatistik $F = 2.17$. 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.

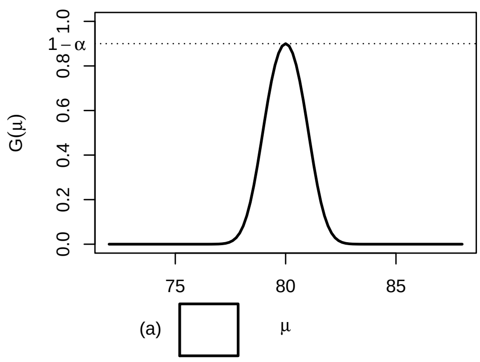


3. Beim zweiseitigen t-Test für den Mittelwert normalverteilter Zufallsvariablen bei unbekannter Varianz führt ein beobachteter Abstand des Stichprobenmittelwerts $\overline{x}$ zum „Sollwert“ μ_0 umso eher zur Ablehnung von H_0, je
   - (a) geringer die Streuung s der Stichprobe und je geringer das Signifikanzniveau α ist. ☐
   - (b) größer die Streuung s der Stichprobe und je geringer das Signifikanzniveau α ist. ☐
   - (c) geringer die Streuung s der Stichprobe und je größer das Signifikanzniveau α ist. ☐
   - (d) größer die Streuung s der Stichprobe und je größer das Signifikanzniveau α ist. ☐

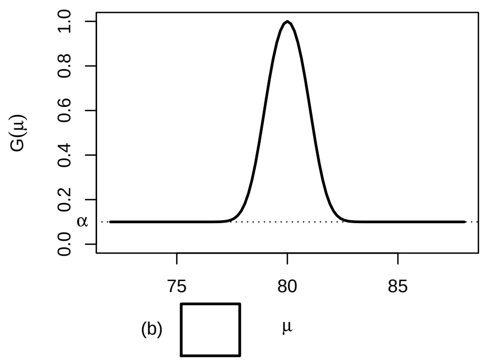
4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{40}$ vom Umfang $n = 40$ zu einer $N(\mu, 4^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu = 80 \qquad \text{gegen} \qquad H_1 : \mu \neq 80$$

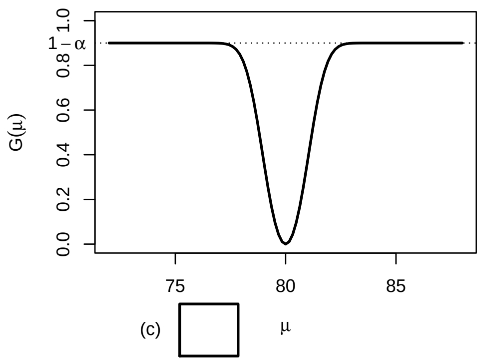
bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.
Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.



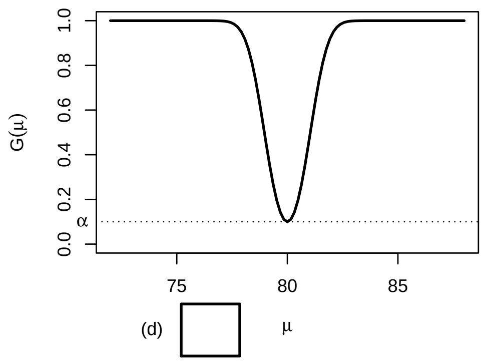
(a) ☐



(b) ☐



(c) ☐



(d) ☐

Aufgabe 3 (6 + 2 = 8 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $a > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|a) = \begin{cases} \dfrac{a^3}{2} \cdot (y+1)^2 \cdot e^{-a\cdot(y+1)} & \text{für } y > -1 \\ 0 & \text{sonst} \end{cases}$$

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) Man kann zeigen, dass $\mathrm{E}(Y) = \frac{3}{a} - 1$ gilt. Bestimmen Sie damit den Schätzer $\widehat{a}_{MM}$ nach der Methode der Momente.

Hinweise:

- *Beachten Sie, dass Sie Teil (b) auch ohne die Bearbeitung von Teil (a) 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.*

Aufgabe 4 (2 + 3 + 3 = 8 Punkte)

Zu $Y \sim N(\mu, \sigma^2)$ liegen die unabhängigen einfachen Stichproben $X_1^A, \ldots, X_{10}^A$ vom Umfang 10 und $X_1^B, \ldots, X_{20}^B$ vom Umfang 20 vor. Mit $\overline{X^A} := \frac{1}{10}\sum_{i=1}^{10} X_i^A$ und $\overline{X^B} := \frac{1}{20}\sum_{i=1}^{20} X_i^B$ werden die Schätzfunktionen

- $\widehat{\mu}_1 = \frac{1}{2} \cdot \overline{X^A} + 2 \cdot \overline{X^B}$,
- $\widehat{\mu}_2 = \frac{1}{2} \cdot \overline{X^A} + \frac{1}{2} \cdot \overline{X^B}$ und
- $\widehat{\mu}_3 = \frac{1}{3} \cdot \overline{X^A} + \frac{2}{3} \cdot \overline{X^B}$

zur Schätzung von μ betrachtet.

(a) Wie sind $\overline{X^A}$ und $\overline{X^B}$ verteilt?

(b) Welche der Schätzfunktionen $\widehat{\mu}_1$, $\widehat{\mu}_2$ und $\widehat{\mu}_3$ sind erwartungstreu für μ? Begründen Sie Ihre Antwort.

(c) Berechnen Sie zu den für μ erwartungstreuen Schätzfunktionen die zugehörige Varianz. Welche dieser Schätzfunktionen würden Sie am ehesten zur Schätzung von μ einsetzen? Begründen Sie Ihre Antwort.

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

Bei der Abfüllung von Sanitärsilikon weiß der Hersteller aus langjähriger Erfahrung, dass die verwendete Maschine eine Standardabweichung von $1.5[ml]$ für die abgefüllte Menge hat. Nach einer routinemäßigen Überprüfung hat der Hersteller den Verdacht, dass die Abfüllanlage im Mittel mehr als die auf dem Produkt ausgezeichneten $310[ml]$ in die Kartuschen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 9 Kartuschen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_9$ als Realisation einer einfachen Stichprobe vom Umfang 9 zur annahmegemäß $N(\mu, 1.5^2[ml^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = 311.183[ml] .$$

(a) 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.

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

(c) Wie groß ist die Wahrscheinlichkeit des Fehlers 2. Art zu dem Test aus Teil (a), falls $\mu = 311.5[ml]$ beträgt?

Aufgabe 6 (10 + 8 = 18 Punkte)

Es werde angenommen, dass die (in Ladezyklen gemessene) Lebensdauer Y^A eines in der laufenden Produktion eingesetzten Akku-Modells normalverteilt sei mit unbekanntem Erwartungswert μ_A und unbekannter Varianz σ_A^2. Es wird erwogen, in der Produktion zukünftig ein alternatives Modell zu verwenden, dessen Lebensdauer Y^B ebenfalls als normalverteilt (mit unbekanntem Erwartungswert μ_B und unbekannter Varianz σ_B^2) angenommen werden kann. Es soll überprüft werden, ob das alternative Modell im Mittel eine höhere Lebensdauer als das aktuell verwendete Modell besitzt.
Aus einem Langzeittest mit $n_A = 11$ Akkus des aktuell verwendeten und $n_B = 9$ Exemplaren des alternativen Modells erhielt man Realisationen jeweils voneinander unabhängiger einfacher Stichproben $X_1^A, \ldots, X_{11}^A$ zu Y^A sowie $X_1^B, \ldots, X_9^B$ zu Y^B und hieraus die zugehörigen Mittelwerte $\overline{x^A} = 1022.27$ bzw. $\overline{x^B} = 1092.67$ sowie die Stichprobenvarianzen $s_{Y^A}^2 = 1745.42$ bzw. $s_{Y^B}^2 = 2511$.

(a) Testen Sie unter der Annahme $\sigma_A^2 = \sigma_B^2$ zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass das alternative Modell im Mittel eine höhere Lebensdauer als das aktuell verwendete Modell besitzt. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.10$, ob die in Teil (a) getroffene Annahme der Varianzgleichheit auf Grundlage der vorhandenen Stichprobeninformation verworfen werden muss. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m,n)$*-Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

Aufgabe 7 (10 + 3 + 2 = 15 Punkte)

In einer bestimmten Multiple-Choice-Aufgabe ist genau eine von vier Antwortmöglichkeiten, die mit „A“, „B“, „C“ bzw. „D“ bezeichnet sind, korrekt. Bei der Korrektur der Klausur stellt der Dozent fest, dass die von den Teilnehmern der Klausur abgegebenen Antworten wie folgt verteilt sind:

Antwort	A	B	C	D
Anteil in Prozent	32	18	22	28

Der Dozent fragt sich, ob man bei dieser Verteilung der abgegebenen Antworten davon ausgehen kann, dass sich die Teilnehmer der Klausur rein zufällig (und voneinander unabhängig) für eine der vier Antworten entschieden haben.

(a) Gehen Sie zunächst davon aus, dass 200 Studierende an der Klausur teilgenommen (und die Aufgabe bearbeitet) haben. Überprüfen Sie unter der Annahme, dass sich die abgegebenen Antworten als einfache Stichprobe auffassen lassen, mit einem geeigneten statistischen Test, ob sich die Prüflinge rein zufällig für eine der vier Antworten entschieden haben oder nicht (Signifikanzniveau $\alpha = 0.05$).

(b) Würde sich das Ergebnis des Tests aus Teil (a) ändern, wenn nur 100 Prüflinge an der Klausur teilgenommen (und die Aufgabe bearbeitet) hätten? Begründen Sie Ihre Antwort! (*Beachten Sie, dass Sie den Test nicht komplett neu durchführen müssen und insbesondere die realisierte Teststatistik relativ leicht aus dem Ergebnis des vorherigen Aufgabenteils gewinnen können!*)

(c) Ab welcher Anzahl von Klausurteilnehmern würde eine Durchführung des bereits in den Teilen (a) und (b) verwendeten Tests dazu führen, davon auszugehen, dass sich die Prüflinge *nicht* rein zufällig für eine der vier Antwortmöglichkeiten entschieden haben?

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

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:

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

Aus vorliegenden Daten zu den Jahren 2007–2013 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
-67.8860  22.6054  -0.4926   0.2221  -8.2959  22.5067  31.3403
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  61.5209   143.0701   0.430  0.68509
x             1.1169     0.1706   6.546  0.00125 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.54 on 5 degrees of freedom
Multiple R-squared:  0.8955,       Adjusted R-squared:  0.8746
F-statistic: 42.85 on 1 and 5 DF,  p-value: 0.001246
```

(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 $\alpha = 0.10$, ob β_1 signifikant positiv ist.

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

(f) Welche Warenausfuhrhöhe (in Milliarden Euro) prognostiziert das Modell für ein Jahr mit einer Wareneinfuhrhöhe von 900 (in Milliarden Euro)?

Aufgabe 9 (6 + 2 + 3 + 5 + 3 + 5 = 24 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 $n = 30$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{30} y_i = 382; \quad \sum_{i=1}^{30} y_i^2 = 5098.139; \quad \sum_{i=1}^{30} x_i = 148.483;$$

$$\sum_{i=1}^{30} x_i^2 = 775.668; \quad \sum_{i=1}^{30} x_i \cdot y_i = 1952.706$$

(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 positiv ist. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

(e) Geben Sie ein symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.90$ für β_1 an.

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

2 Schließende Statistik SS 2015

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. Seien $X_1, \dots, X_n$ sowie Y normalverteilt mit $$\mathrm{E}(X_1) = \mathrm{E}(X_2) = \ldots = \mathrm{E}(X_n) = \mathrm{E}(Y)\ .$$ Dann ist $X_1, \dots, X_n$ stets eine einfache Stichprobe vom Umfang n zu Y.	☐	☐
2. Ist eine Familie von Schätzfunktionen $T_n, n \in \mathbb{N}$, konsistent im quadratischen Mittel für einen Parameter $\lambda \in \mathbb{R}$, so ist jede dieser Schätzfunktionen auch effizient in der Klasse der für λ erwartungstreuen Schätzfunktionen.	☐	☐
3. Setzt man den aus einer Realisation $x_1, \dots, x_n$ einer einfachen Stichprobe nach der Momentenmethode erhaltenen Parameterschätzwert in die zugehörige Likelihoodfunktion ein, so ist es möglich, dass man dabei den Wert 0 erhält.	☐	☐
4. Wird bei einem statistischen Hypothesentest die Nullhypothese beibehalten, obwohl sie falsch ist, handelt es sich um einen Fehler 2. Art.	☐	☐
5. Gilt bei der Anwendung eines statistischen Tests für den p-Wert p zur Teststatistik T die Beziehung $p < \alpha$, so liegt maximal mit Wahrscheinlichkeit p ein Fehler 1. Art vor.	☐	☐
6. Nimmt ein zweiseitiger Gauß-Test mit $H_0 : \mu = \mu_0$ die Nullhypothese H_0 zum Signifikanzniveau $\alpha = 0.05$ an, so entscheiden auch die beiden einseitigen Tests mit $H_0 : \mu \leq \mu_0$ bzw. $H_0 : \mu \geq \mu_0$ zum Signifikanzniveau $\alpha = 0.05$ stets zu Gunsten der Nullhypothese H_0.	☐	☐
7. Die einfache Varianzanalyse mit 2 Faktorstufen sowie der zweiseitige 2-Stichproben-t-Test liefern unter den üblichen Annahmen (unabhängig normalverteilte Zufallsvariablen mit identischen Varianzen) in übereinstimmenden Anwendungssituationen auch stets übereinstimmende Ergebnisse.	☐	☐
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 die Kleinst-Quadrate-Schätzfunktionen $\widehat{\beta}_1$ und $\widehat{\beta}_2$ linear in y_i.	☐	☐

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. Konfidenzintervalle für den Erwartungswert einer normalverteilten Zufallsvariablen mit bekannter Varianz sind umso breiter,

   (a) je kleiner der Stichprobenumfang und je kleiner das Konfidenzniveau $1 - \alpha$ ist.

   (b) je kleiner der Stichprobenumfang und je größer das Konfidenzniveau $1 - \alpha$ ist.

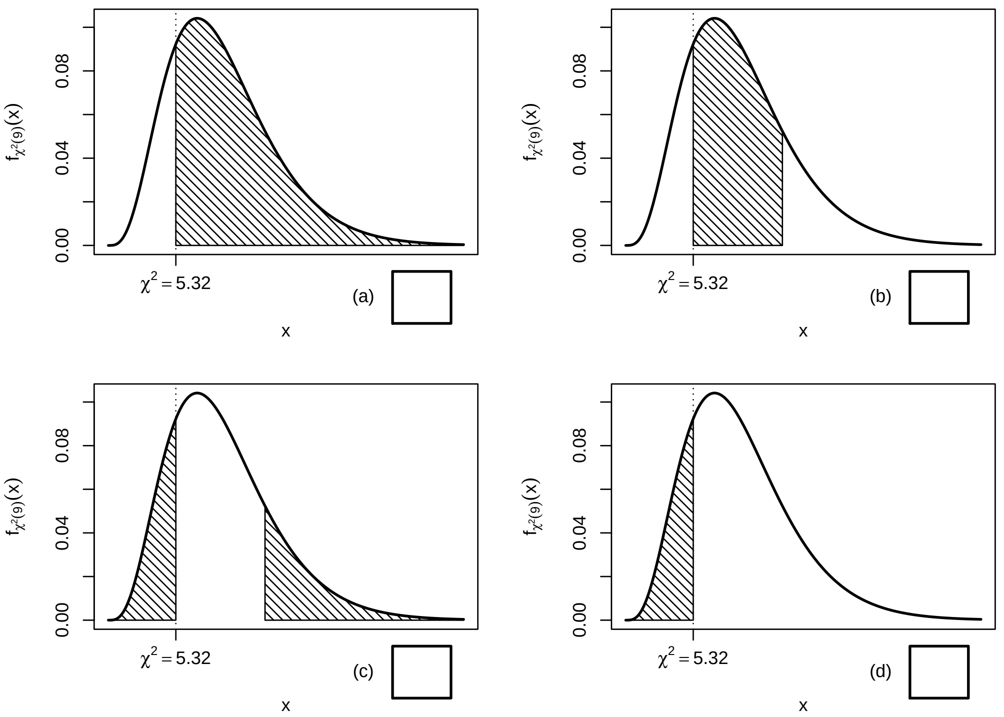
   (c) je größer der Stichprobenumfang und je kleiner das Konfidenzniveau $1 - \alpha$ ist.

   (d) je größer der Stichprobenumfang und je größer das Konfidenzniveau $1 - \alpha$ ist.

2. Sei $X_1, \ldots, X_{10}$ 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 $n = 10$ soll

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

   mit einem Chi-Quadrat-Test getestet werden. Als realisierte Teststatistik erhält man $\chi^2 = 5.32$. 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. 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 rechtsseitige 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) $t \in [-t_{n-1,1-\alpha}, t_{n-1,1-\alpha}]$ ☐

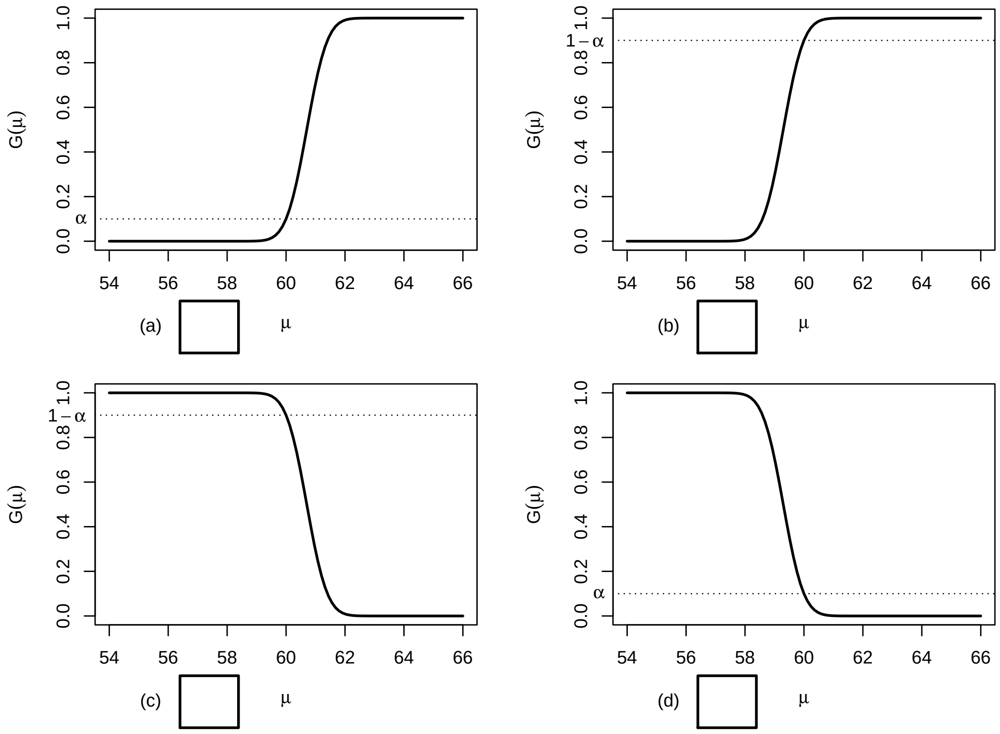
   (c) $t \in (t_{n-1,1-\alpha}, t_{n-1,1-\frac{\alpha}{2}}]$ ☐

   (d) $t \in (t_{n-1,1-\frac{\alpha}{2}}, \infty)$ ☐

4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{30}$ vom Umfang $n = 30$ zu einer $N(\mu, 3^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu \geq 60 \qquad \text{gegen} \qquad H_1 : \mu < 60$$

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



Aufgabe 3 (2 + 4 = 6 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $\lambda > 0$ eine stetige Gleichverteilung auf dem Intervall $\left[0, \lambda\right]$. Der quadrierte Parameter λ^2 soll auf Grundlage einer einfachen Stichprobe $X_1, \ldots, X_n$ vom Umfang n geschätzt werden.

(a) Bekanntlich haben auf dem Intervall $[a, b]$ (mit $a < b$) stetig gleichverteilte Zufallsvariablen den Erwartungswert $\dfrac{a+b}{2}$ sowie die Varianz $\dfrac{(b-a)^2}{12}$. Bestimmen Sie damit den Erwartungswert und die Varianz von Y in Abhängigkeit des unbekannten Parameters λ.

(b) Überprüfen Sie mit Hilfe von Teil (a), ob

$$\widehat{\lambda^2} := 3 \cdot \overline{X^2} = \frac{3}{n} \sum_{i=1}^{n} X_i^2$$

erwartungstreu für λ^2 ist.

Aufgabe 4 (6 + 2 = 8 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $a > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|a) = \begin{cases} a^2 \cdot (y+2) \cdot e^{-a\cdot(y+2)} & \text{für } y > -2 \\ 0 & \text{sonst} \end{cases}$$

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) Man kann zeigen, dass $\mathrm{E}(Y) = \frac{2}{a} - 2$ gilt. Bestimmen Sie damit den Schätzer $\widehat{a}_{MM}$ nach der Methode der Momente.

Hinweise:

- *Beachten Sie, dass Sie Teil (b) auch ohne die Bearbeitung von Teil (a) 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.*

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

Bei der Abfüllung von Druckgaspatronen weiß der Hersteller aus langjähriger Erfahrung, dass die verwendete Maschine eine Standardabweichung von $2[g]$ 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 $330[g]$ in die Patronen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 20 Patronen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{20}$ als Realisation einer einfachen Stichprobe vom Umfang 20 zur annahmegemäß $N(\mu, 2^2[g^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = 329.141[g] \; .$$

(a) 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.

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

(c) Wie groß ist die Wahrscheinlichkeit des Fehlers 2. Art zu dem Test aus Teil (a), falls $\mu = 329[g]$ beträgt?

Aufgabe 6 (11 + 9 + 8 = 28 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 9 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	9
Marke A x_i^A	269	289	278	300	308	311	263	313	288
Marke B x_i^B	276	282	254	290	303	287	269	283	275

(a) 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 höhere Aufnahmeanzahl 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 Aufnahmeanzahlen zu den jeweiligen Digitalkameramodellen verloren gegangen ist. Um die Situation zu retten, nehme man weiter an, dass mit $X_1^A, \ldots, X_9^A$ und $X_1^B, \ldots, X_9^B$ 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 Batteriemarke A im Vergleich zu Batteriemarke B eine höhere Aufnahmeanzahl ermöglicht. Verwenden Sie hierzu die Stichprobenmittelwerte $\overline{x^A} = 291$ bzw. $\overline{x^B} = 279.89$ sowie die Stichprobenvarianzen $s^2_{Y^A} = 338$ bzw. $s^2_{Y^B} = 191.11$. Fassen Sie das Ergebnis des Tests auch in einem Antwortsatz zusammen.

(c) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.10$, ob die in Teil (b) getroffene Annahme der Varianzgleichheit auf Grundlage der vorhandenen Stichprobeninformation verworfen werden muss. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (c) den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m, n)$*-Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

Aufgabe 7 (11 Punkte)

Mit einem Chi-Quadrat-Anpassungstest soll zum Signifikanzniveau $\alpha = 0.05$ überprüft werden, ob man bei einem beobachteten Stichprobenergebnis von der Realisation einer einfachen Stichprobe vom Umfang $n = 100$ zu einer Geom(0.35)-verteilten Zufallsvariablen ausgehen kann. Die Stichprobeninformation liege in Form der folgenden Häufigkeitsverteilung vor:

a_i	0	1	2	≥ 3
n_i	26	19	13	42

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

Hinweise:

- *Die geometrische Verteilung mit Parameter* $p = 0.35$ *hat den Träger* $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$ *und die Wahrscheinlichkeitsfunktion*

$$p_{\text{Geom}(0.35)} : \mathbb{N}_0 \to [0, 1]; p_{\text{Geom}(0.35)}(i) = (1 - 0.35)^i \cdot 0.35 \,.$$

- *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

Aufgabe 8 (1 + 1 + 1 + 1 + 1 + 1 + 1 = 7 Punkte)

Zur Erklärung der stetigen Wochenrenditen der Beiersdorf-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 der jüngeren Vergangenheit wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
    Min      1Q  Median      3Q     Max
-3.6452 -2.1621  0.1126  1.4974  6.3684

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.07654    0.54402  -0.141   0.8893
x            0.59654    0.33326   1.790   0.0861 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.568 on 24 degrees of freedom
Multiple R-squared:  0.1178,	Adjusted R-squared:  0.08102
F-statistic: 3.204 on 1 and 24 DF,  p-value: 0.08608
```

(a) Wie viele Wochenrenditen 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 stetigen Wochenrenditen der Beiersdorf-Aktie wird durch das lineare Modell erklärt?

(e) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.01$, ob β_1 signifikant von Null verschieden ist.

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

(g) Welche stetige Wochenrendite der Beiersdorf-Aktie prognostiziert das Modell in einer Woche mit stetiger DAX-Rendite von 0.6 (in Prozent)?

Aufgabe 9 (6 + 2 + 3 + 3 + 5 = 19 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:

$$\sum_{i=1}^{20} y_i = 485.086; \quad \sum_{i=1}^{20} y_i^2 = 12174.045; \quad \sum_{i=1}^{20} x_i = 108.964;$$

$$\sum_{i=1}^{20} x_i^2 = 650.739; \quad \sum_{i=1}^{20} x_i \cdot y_i = 2789.235$$

(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 symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_1 an.

(e) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für y_0 gegeben $x_0 = 4.5$ an.

3 Schließende Statistik WS 2015/16

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. Zufallsstichproben zu einer normalverteilen Zufallsvariablen Y sind stets einfache (Zufalls-)Stichproben.	☐	☐
2. Setzt man den aus einer Realisation $x_1, \dots, x_n$ einer einfachen Stichprobe nach der Momentenmethode erhaltenen Parameterschätzwert in die zugehörige Likelihoodfunktion ein, so ist der erhaltene Wert stets kleiner als das Maximum der Likelihoodfunktion, welches bei Einsetzen des nach der ML-Methode erhaltenen Schätzwerts angenommen wird.	☐	☐
3. Im quadratischen Mittel für einen Parameter θ konsistente Schätzfunktionen sind nie erwartungstreu für θ.	☐	☐
4. Je größer der p-Wert beim rechtsseitigen Gauß-Test für den Erwartungswert bei bekannter Varianz ist, umso kleiner ist die Fehlerwahrscheinlichkeit 1. Art.	☐	☐
5. Gilt für den p-Wert bei Anwendung eines linksseitigen t-Tests für den Erwartungswert $p \leq 0.05$, so ist die Differenz $\mu - \mu_0$ zwischen dem wahren Erwartungswert μ und μ_0 signifikant niedriger als 0.05.	☐	☐
6. Nimmt ein zweiseitiger Gauß-Test mit $H_0 : \mu = \mu_0$ die Nullhypothese H_0 zum Signifikanzniveau $\alpha = 0.10$ an, so entscheidet höchstens einer der beiden einseitigen Tests mit $H_0 : \mu \leq \mu_0$ bzw. $H_0 : \mu \geq \mu_0$ zum Signifikanzniveau $\alpha = 0.05$ ebenfalls zu Gunsten der Nullhypothese H_0.	☐	☐
7. Die Teststatistik der einfachen Varianzanalyse kann wegen der Gültigkeit der Streuungszerlegung $SS = SB + SW$ auch berechnet werden, wenn als Stichprobeninformation (neben den Einzelstichprobenumfängen $n_1, \dots, n_k$) nur die Stichprobenvarianzan $S_1^2, \dots, S_k^2$ aus den Einzelstichproben sowie S^2 aus der Gesamtstichprobe zur Verfügung stehen.	☐	☐
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 die Kleinst-Quadrate-Schätzfunktionen $\widehat{\beta}_1$ und $\widehat{\beta}_2$ stets normalverteilt.	☐	☐

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. Es sei $X_1, \dots, X_{16}$ eine einfache Stichprobe vom Umfang 16 zu Y mit $Y \sim N(100, 16^2)$. Dann gilt für die Verteilung von $\overline{X} = \frac{1}{16} \sum_{i=1}^{16} X_i$:

   (a) $\overline{X} \sim N\left(100, 64^2\right)$ ☐

   (b) $\overline{X} \sim N\left(100, 1^2\right)$ ☐

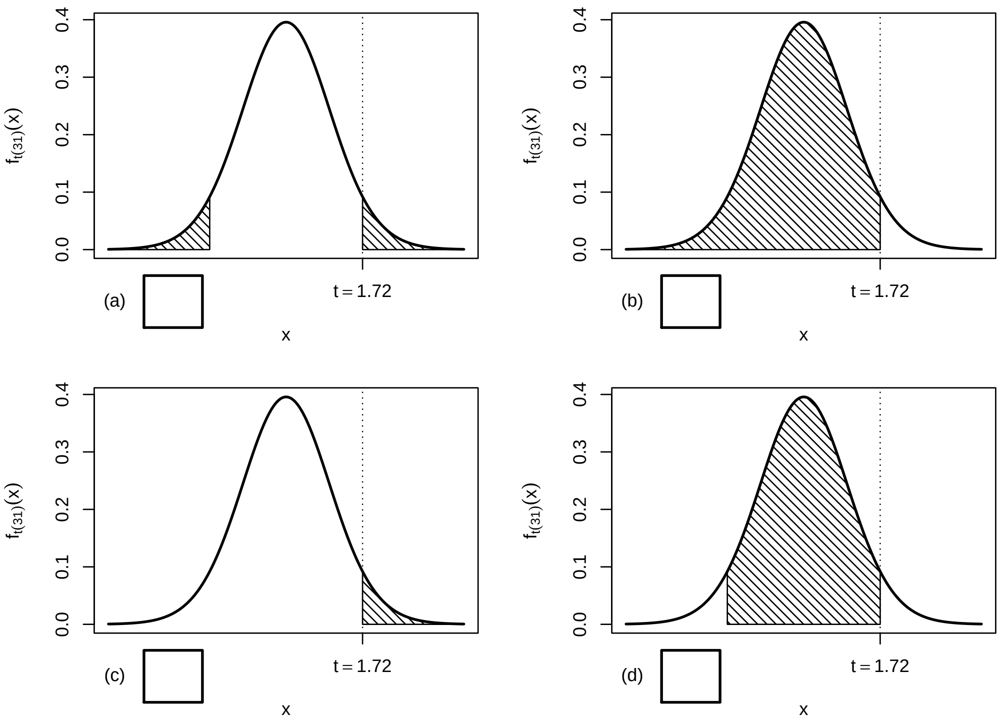
   (c) $\overline{X} \sim N\left(100, 16^2\right)$ ☐

   (d) $\overline{X} \sim N\left(100, 4^2\right)$ ☐

2. Sei $X_1, \dots, X_{32}$ 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 $n = 32$ soll

$$H_0 : \mu = \mu_0 = 20 \qquad \text{gegen} \qquad H_1 : \mu \neq \mu_0 = 20$$

   mit einem t-Test getestet werden. Als realisierte Teststatistik erhält man $t = 1.72$. 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. Wird die Teststatistik der einfachen Varianzanalyse als Quotient mit dem Zähler $SB/(k-1)$ und dem Nenner $SW/(n-k)$ notiert und bezeichnet σ^2 die Varianz der Zufallsvariablen $Y_1, \ldots, Y_k$, so

   (a) sind Zähler und Nenner stets sinnvolle Schätzer für σ^2. ☐

   (b) ist der Zähler nur unter H_0, der Nenner stets ein sinnvoller Schätzer für σ^2. ☐

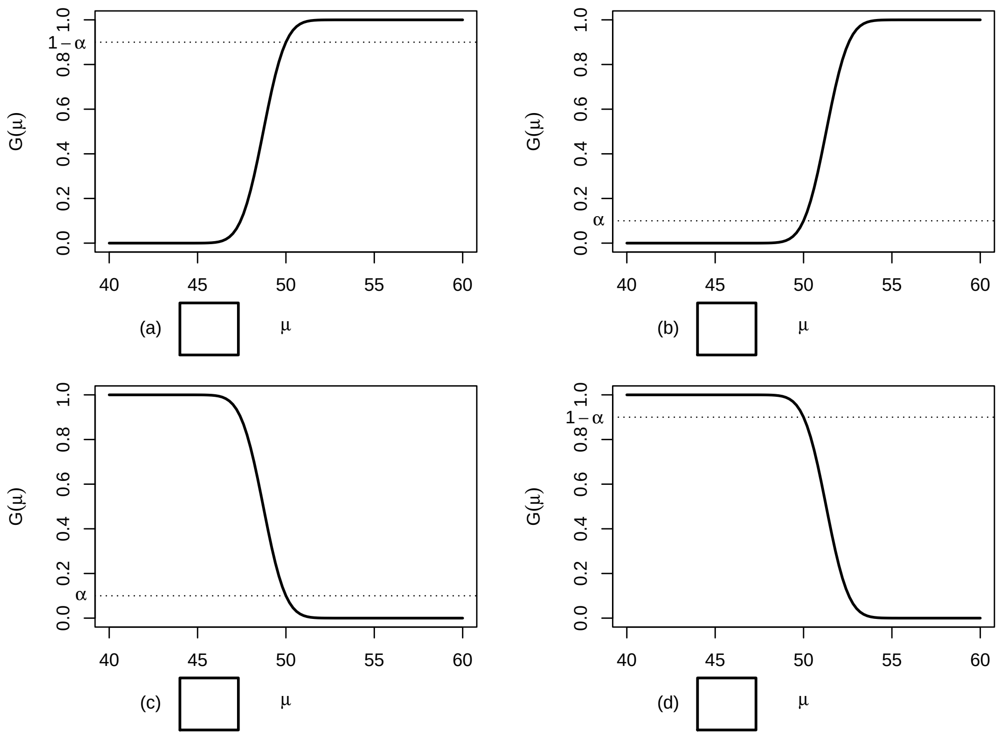
   (c) ist der Zähler stets, der Nenner nur unter H_0 ein sinnvoller Schätzer für σ^2. ☐

   (d) sind Zähler und Nenner nur unter H_0 sinnvolle Schätzer für σ^2. ☐

4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{25}$ vom Umfang $n = 25$ zu einer $N(\mu, 5^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu \leq 50 \qquad \text{gegen} \qquad H_1 : \mu > 50$$

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



Aufgabe 3 (6 Punkte)

Wird die Alternativverteilung mit Parameter p als (allgemeine) diskrete Verteilung mit den beiden Trägerpunkten $a_1 = 0$ (mit Punktwahrscheinlichkeit $1-p$) und $a_2 = 1$ (mit Punktwahrscheinlichkeit p) aufgefasst, so kann statt des approximativen Gauß-Tests mit der Nullhypothese $H_0 : p = p_0$ und der Alternative $H_1 : p \neq p_0$ offensichtlich auch ein Chi-Quadrat-Anpassungstests mit $p^0 = (1-p_0, p_0)$ durchgeführt werden.
Zeigen Sie, dass zwischen den beiden Teststatistiken χ^2 des Chi-Quadrat-Anpassungstests

und N des approximativen Gauß-Tests die Beziehung

$$\chi^2 = \sum_{i=1}^{2} \frac{(n_i - np_i^0)^2}{np_i^0} \stackrel{!}{=} \frac{(\widehat{p} - p_0)^2}{p_0 \cdot (1 - p_0)} \cdot n = N^2$$

gilt.

Hinweis:
Sind in einer einfachen Stichprobe vom Umfang n insgesamt n_1 Misserfolge („Nullen“) und n_2 Erfolge („Einsen“) enthalten, so gilt offensichtlich $\widehat{p} = \frac{n_2}{n}$ sowie $n_2 = n - n_1$.

Aufgabe 4 (3 + 1 + 6 = 10 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $b > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|b) = \begin{cases} 24 \cdot b^{-3} \cdot y^2 & \text{für } 0 \leq y \leq \frac{1}{2} \cdot b \\ 0 & \text{sonst} \end{cases}$$

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

(a) Zeigen Sie, dass $\mathrm{E}(Y) = \frac{3}{8} \cdot b$ gilt.

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

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

Hinweise:

- *Beachten Sie, dass Sie die Teile (b) und (c) auch ohne die Bearbeitung von Teil (a) 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.*

Aufgabe 5 (7 + 2 + 3 = 12 Punkte)

Bei der Abfüllung von Spaghetti weiß der Hersteller aus langjähriger Erfahrung, dass die verwendete Maschine eine Standardabweichung von $2[g]$ 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 $500[g]$ in die Packungen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 25 Packungen entnommen, deren gemessene Füllmengen $x_1, \dots, x_{25}$ als Realisation einer einfachen Stichprobe vom Umfang 25 zur annahmegemäß $N(\mu, 2^2[g^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = 499.315[g] \ .$$

(a) 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.

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

(c) Geben Sie auf Basis der Stichprobenrealisation ein zweiseitiges Konfidenzintervall für die mittlere Abfüllmenge zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.99$(!) an.

Aufgabe 6 (9 + 7 = 16 Punkte)

Ein Hersteller von Metallwaren produziert unter anderem Schrauben mit einer Soll-Länge von 7 [cm]. Es soll angenommen werden, dass die Länge der hergestellten Schrauben gemäß einer Normalverteilung mit unbekanntem Erwartungswert und unbekannter Varianz schwankt. Zur Überprüfung der Qualität der produzierten Schrauben werden 9 Schrauben aus dem laufenden Produktionsprozess entnommen. Die gemessenen Längen

$$7.018, 7.128, 6.976, 7.063, 7.034, 6.912, 7.002, 7.076, 6.991$$

seien als Realisation einer einfachen Stichprobe zur normalverteilten Grundgesamtheit aufzufassen. Aus dieser Realisation wurde bereits $s^2 = 0.003946$ berechnet.

(a) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die tatsächliche mittlere Länge der produzierten Schrauben von der angegebenen Soll-Länge von 7 [cm] abweicht. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die Varianz der Länge der produzierten Schrauben im Vergleich zur vom Hersteller angegebenen Toleranz $\sigma_0^2 = 0.0016$ zu groß ist. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) 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
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

Aufgabe 7 (10 Punkte)

Zwei unterschiedlichen Gruppen mit 75 (Gruppe A) bzw. 67 (Gruppe B) Blutdruckpatienten wird jeweils ein spezieller Blutdrucksenker verabreicht. Nach einer festgelegten Zeit werden dann alle Blutdruckpatienten gefragt, ob durch das verabreichte Medikament eine blutdrucksenkende Wirkung eingetreten ist. In der Gruppe der Blutdruckpatienten, denen Blutdrucksenker A verabreicht wurde, beantworten 45 Personen diese Frage positiv, in der zu Blutdrucksenker B gehörigen Gruppe 51 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 Blutdrucksenker B besser wirkt als Blutdrucksenker A (bezogen auf die Erfolgswahrscheinlichkeit für eine blutdrucksenkende Wirkung). Formulieren Sie das Ergebnis auch in Form eines Antwortsatzes.

Aufgabe 8 (13 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 2014/15 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	$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{j,i}$	$\sum_{i=1}^{n_j} x_{j,i}^2$
1	89	83.517	656703.00
2	17	89.235	140037.00
3	55	100.045	563757.25

Für die Durchführung der einfachen Varianzanalyse wurde hieraus bereits die Größe $SW = 53850.143$ berechnet.
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 $X_{j,i} (1 \le j \le 3, 1 \le i \le n_j)$ 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.

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

$n\backslash m$	1	2	3	4	5	155	156	157	158	159
1	161.448	199.500	215.707	224.583	230.162	253.492	253.497	253.503	253.508	253.513
2	18.513	19.000	19.164	19.247	19.296	19.489	19.489	19.489	19.489	19.489
3	10.128	9.552	9.277	9.117	9.013	8.544	8.544	8.544	8.544	8.544
4	7.709	6.944	6.591	6.388	6.256	5.651	5.651	5.651	5.651	5.651
5	6.608	5.786	5.409	5.192	5.050	4.391	4.391	4.391	4.390	4.390
155	3.902	3.054	2.663	2.430	2.273	1.303	1.303	1.303	1.302	1.302
156	3.902	3.054	2.663	2.430	2.272	1.303	1.302	1.302	1.301	1.301
157	3.901	3.054	2.662	2.429	2.272	1.302	1.302	1.301	1.301	1.300
158	3.901	3.053	2.662	2.429	2.271	1.302	1.301	1.301	1.300	1.300
159	3.901	3.053	2.661	2.429	2.271	1.301	1.300	1.300	1.300	1.299

Aufgabe 9 ($1 + 1 + 1 + 1 + 1 + 1 = 6$ Punkte)

Zur Erklärung der Höhe der laufenden Grundmittel an deutschen Hochschulen y_i (in Millionen Euro) durch die Anzahl von Studierenden an deutschen Hochschulen x_i (in Millionen) unterstellt man die Gültigkeit eines Zusammenhangs im Sinne des folgenden linearen Modells:

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

Aus vorliegenden Daten zu den Jahren 2008–2013 wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
       1        2        3        4        5        6
-223.594   -8.204  222.692    7.144   76.020  -74.059

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3801.3      950.5   3.999 0.016138 *
x             5358.6      449.1  11.933 0.000283 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 166.6 on 4 degrees of freedom
Multiple R-squared:  0.9727,        Adjusted R-squared:  0.9658
F-statistic: 142.4 on 1 and 4 DF,  p-value: 0.0002826
```

(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 laufenden Grundmittel an deutschen Hochschulen wird durch das lineare Modell erklärt?

(d) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.01$, ob β_1 signifikant positiv ist.

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

(f) Welche Höhe der laufenden Grundmittel (in Millionen Euro) prognostiziert das Modell für ein Jahr mit einer Studierendenanzahl von 2.4 (in Millionen)?

Aufgabe 10 (6 + 2 + 3 + 3 + 5 = 19 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 $n = 20$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{20} y_i = 65.443; \quad \sum_{i=1}^{20} y_i^2 = 424.152; \quad \sum_{i=1}^{20} x_i = 121.795;$$

$$\sum_{i=1}^{20} x_i^2 = 779.684; \quad \sum_{i=1}^{20} x_i \cdot y_i = 336.339$$

(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 symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_1 an.

(e) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.90$ für y_0 gegeben $x_0 = 4$ an.

4 Schließende Statistik SS 2016

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 einer normalverteilten Zufallsvariablen Y. Dann sind auch $X_1, \dots, X_n$ stets normalverteilt.	☐	☐
2. Ist eine Schätzfunktion effizient in einer Klasse erwartungstreuer Schätzfunktionen, so hat keine der Schätzfunktionen in dieser Klasse eine größere Varianz als die effiziente Schätzfunktion selbst.	☐	☐
3. Zur Schätzung des Parameters $\theta \in \mathbb{R}$ seien für $n \in \mathbb{N}$ Schätzfunktionen T_n gegeben mit den Eigenschaften $\mathrm{E}(T_n) = \theta$ und $\mathrm{Var}(T_n) = 2 + \frac{2}{n}$ für alle $n \in \mathbb{N}$. Dann ist die Folge T_n von Schätzfunktionen für θ konsistent im quadratischen Mittel.	☐	☐
4. Ist p der p-Wert eines zweiseitigen Chi-Quadrat-Tests für die Varianz einer normalverteilten Zufallsvariablen Y mit der Nullhypothese $\sigma^2 = 4$, so weicht die tatsächliche Varianz der Zufallsvariablen Y (betragsmäßig) um mindestens p von 4 ab.	☐	☐
5. Gilt für den p-Wert bei Anwendung eines statistischen Hypothesentests zum Signifikanzniveau α die Beziehung $p < \alpha$, so ist die Nullhypothese des Tests nur mit einer Wahrscheinlichkeit von p zutreffend.	☐	☐
6. Bei einem rechtsseitigen Gauß-Test von $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$ kann die Fehlerwahrscheinlichkeit 2. Art für $\mu > \mu_0$ durch $1 - G(\mu)$ berechnet werden, wobei $G(\mu)$ die Gütefunktion des Tests bezeichnet.	☐	☐
7. Mit der einfachen Varianzanalyse kann unter der Annahme, dass sich die Erwartungswerte in den einzelnen Faktorstufen nicht unterscheiden, überprüft werden, ob auch die Varianzen in den Faktorstufen vollständig übereinstimmen.	☐	☐
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 zur Vertrauenswahrscheinlichkeit $1 - \alpha$ für y_0 stets größer als die entsprechenden Prognoseintervalle für $\mathrm{E}(y_0)$.	☐	☐

Aufgabe 2 (12 Punkte)

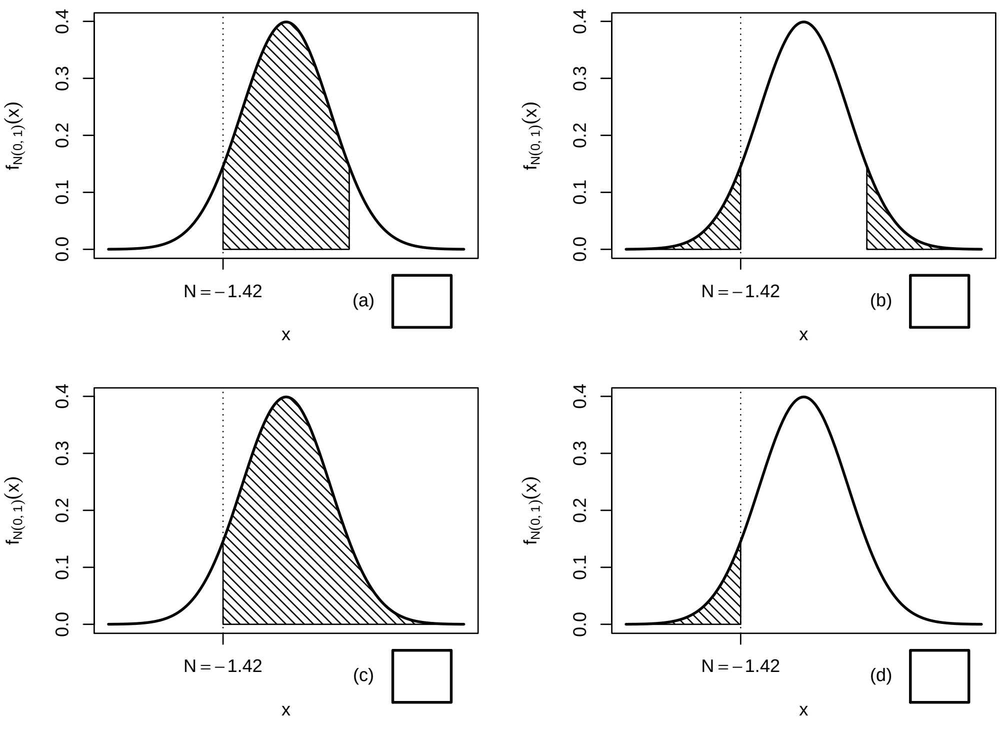
Markieren Sie jeweils die korrekte Antwort mit einem Kreuz im betreffenden Kästchen. Es ist jeweils genau ein Kreuz korrekt.
Richtige Antworten geben 3 Punkte, falsche Antworten und nicht bearbeitete Aufgabenteile 0 Punkte (Aufgabenteile mit mehr als einem Kreuz zählen als nicht bearbeitet!).

1. Mit einem Chi-Quadrat-Anpassungstest soll anhand einer einfachen Stichprobe vom Umfang $n = 100$ überprüft werden, ob die Grundgesamtheit normalverteilt ist. Nach Wahl einer geeigneten Klassierung aus 5 Klassen werden dazu zunächst die beiden unbekannten Parameter der Normalverteilung durch eine ML-Schätzung aus den klassierten Daten ermittelt. Damit ist zur Konstruktion des Ablehnungsbereiches die folgende Verteilung zu verwenden:
   (a) χ^2-Verteilung mit 198 Freiheitsgraden
   (b) χ^2-Verteilung mit 197 Freiheitsgraden
   (c) χ^2-Verteilung mit 3 Freiheitsgraden
   (d) χ^2-Verteilung mit 2 Freiheitsgraden

2. Sei $X_1, \ldots, X_{25}$ eine einfache Stichprobe zu einer $N(\mu, \sigma_0^2)$-verteilten Zufallsvariablen Y mit unbekanntem Parameter μ und bekanntem $\sigma_0^2 = 5^2$. Auf der Grundlage einer Stichprobenrealisation zu dieser einfachen Stichprobe vom Umfang $n = 25$ soll

$$H_0 : \mu \leq \mu_0 = 50 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0 = 50$$

   mit einem Gauß-Test getestet werden. Als realisierte Teststatistik erhält man $N = -1.42$. 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.

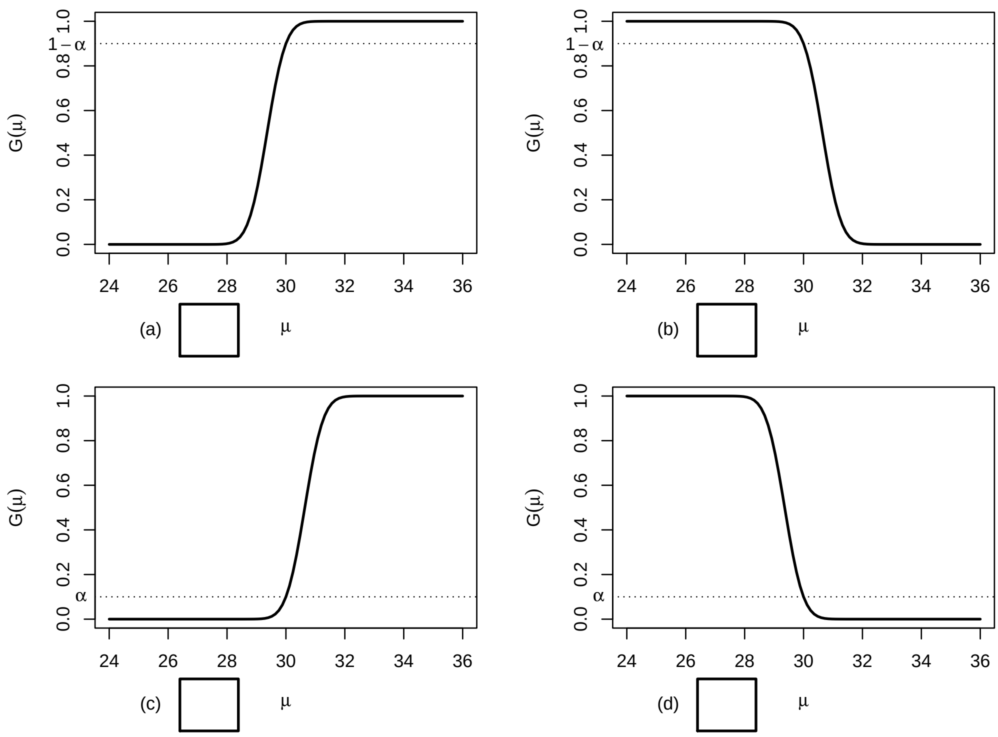


3. Als p-Wert zur realisierten Teststatistik eines zweiseitigen Gauß-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz erhält man $p = 0.04733$. Dann gilt:
   (a) Die Nullhypothese ist bei einem Test zum Niveau $\alpha = 0.10$ abzulehnen, bei einem Test zum Niveau $\alpha = 0.01$ jedoch nicht abzulehnen. ☐
   (b) Die Nullhypothese ist bei einem Test zum Niveau $\alpha = 0.01$ abzulehnen, bei einem Test zum Niveau $\alpha = 0.10$ jedoch nicht abzulehnen. ☐
   (c) Die Nullhypothese ist sowohl bei einem Test zum Niveau $\alpha = 0.01$ als auch bei einem Test zum Niveau $\alpha = 0.10$ abzulehnen ☐
   (d) Die Nullhypothese ist weder bei einem Test zum Niveau $\alpha = 0.01$ noch bei einem Test zum Niveau $\alpha = 0.10$ abzulehnen. ☐

4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{36}$ vom Umfang $n = 36$ 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.



Aufgabe 3 (4 Punkte)

Für $\lambda > 0$ sei $Y \sim \text{Exp}(\lambda)$, es gilt also insbesondere $\text{E}(Y) = \frac{1}{\lambda}$ sowie $\text{Var}(Y) = \frac{1}{\lambda^2}$. $X_1, \ldots, X_n$ sei für $n \in \mathbb{N}$ eine einfache Stichprobe vom Umfang n zu Y. Untersuchen Sie, ob die Schätzfunktionen

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

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

Aufgabe 4 (3 + 1 + 6 = 10 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $\lambda > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|\lambda) = \begin{cases} \frac{3}{2} \cdot \lambda^{-\frac{3}{2}} \cdot y^{\frac{1}{2}} & \text{für } 0 \leq y \leq \lambda \\ 0 & \text{sonst} \end{cases}$$

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

(a) Zeigen Sie, dass $\mathrm{E}(Y) = \frac{3}{5} \cdot \lambda$ gilt.

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

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

Hinweise:

- *Beachten Sie, dass Sie die Teile (b) und (c) auch ohne die Bearbeitung von Teil (a) 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.*

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

Bei der Abfüllung von Kaffeekapseln weiß der Hersteller aus langjähriger Erfahrung, dass die verwendete Maschine eine Standardabweichung von $0.2[g]$ für die abgefüllte Menge hat. Nach einer routinemäßigen Überprüfung hat der Hersteller den Verdacht, dass die Abfüllanlage im Mittel eine andere Menge als die auf dem Produkt ausgezeichneten $7.5[g]$ in die Kapseln einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 15 Kapseln entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{15}$ als Realisation einer einfachen Stichprobe vom Umfang 15 zur annahmegemäß $N(\mu, 0.2^2[g^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = 7.369[g] .$$

(a) Testen Sie zum Signifikanzniveau $\alpha = 0.01$ **(!)**, 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 zu dem Test aus Teil (a), falls $\mu = 7.45[g]$ beträgt?

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

Aufgabe 6 (10 + 8 = 18 Punkte)

Es werde angenommen, dass die (in Kilometern gemessene) Laufleistung Y^A eines aktuell von einer Autovermietung eingesetzten Reifenmodells normalverteilt sei mit unbekanntem Erwartungswert μ_A und unbekannter Varianz σ_A^2. Der Autovermieter erwägt, einen Teil der Fahrzeugflotte in Zukunft mit einem alternativen Reifenmodell auszurüsten, dessen Laufleistung Y^B ebenfalls als normalverteilt (mit unbekanntem Erwartungswert μ_B und unbekannter Varianz σ_B^2) angenommen werden kann. Es soll überprüft werden, ob das alternative Modell im Mittel eine höhere Laufleistung als das aktuell verwendete Modell besitzt.
Aus einem Langzeittest mit $n_A = 12$ Reifensätzen des aktuell verwendeten und $n_B = 15$ Sätzen des alternativen Modells erhielt man Realisationen jeweils voneinander unabhängiger einfacher Stichproben $X_1^A, \ldots, X_{12}^A$ zu Y^A sowie $X_1^B, \ldots, X_{15}^B$ zu Y^B und hieraus die zugehörigen Mittelwerte $\overline{x^A} = 51374$ bzw. $\overline{x^B} = 55882$ sowie die Stichproben**standardabweichungen** $s_{Y^A} = 4763$ bzw. $s_{Y^B} = 5582$.

(a) Testen Sie unter der Annahme $\sigma_A^2 = \sigma_B^2$ zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass das alternative Modell im Mittel eine höhere Laufleistung als das aktuell verwendete Modell besitzt. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.10$, ob die in Teil (a) getroffene Annahme der Varianzgleichheit auf Grundlage der vorhandenen Stichprobeninformation verworfen werden muss. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m, n)$*-Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n \backslash m$	11	12	13	14	15	16	17	18	19	20
11	2.818	2.788	2.761	2.739	2.719	2.701	2.685	2.671	2.658	2.646
12	2.717	2.687	2.660	2.637	2.617	2.599	2.583	2.568	2.555	2.544
13	2.635	2.604	2.577	2.554	2.533	2.515	2.499	2.484	2.471	2.459
14	2.565	2.534	2.507	2.484	2.463	2.445	2.428	2.413	2.400	2.388
15	2.507	2.475	2.448	2.424	2.403	2.385	2.368	2.353	2.340	2.328
16	2.456	2.425	2.397	2.373	2.352	2.333	2.317	2.302	2.288	2.276
17	2.413	2.381	2.353	2.329	2.308	2.289	2.272	2.257	2.243	2.230
18	2.374	2.342	2.314	2.290	2.269	2.250	2.233	2.217	2.203	2.191
19	2.340	2.308	2.280	2.256	2.234	2.215	2.198	2.182	2.168	2.155
20	2.310	2.278	2.250	2.225	2.203	2.184	2.167	2.151	2.137	2.124

Aufgabe 7 (16 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 230 Teilnehmer die folgende Tabelle zusammengestellt:

	0 Blätter bearbeitet	1 Blatt bearbeitet	2 Blätter bearbeitet
bestanden	98	35	62
nicht bestanden	29	4	2

Überprüfen Sie anhand dieses Datenmaterials zum Signifikanzniveau $\alpha = 0.01$ **(!)**, 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

Aufgabe 8 (1 + 1 + 1 + 1 + 1 + 1 + 1 = 7 Punkte)

Zur Erklärung des Geburtsgewichts y_i (in Gramm) durch den per Ultraschall gemessenen Bauchdurchmesser des Fötus unmittelbar vor der Geburt x_i (in Millimetern) unterstellt man die Gültigkeit eines Zusammenhangs im Sinne des folgenden linearen Modells:

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

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

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

Residuals:
    Min      1Q  Median      3Q     Max
-679.17 -224.10  -71.86  119.71 1278.14

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2646.19    1676.46  -1.578   0.1489
x              54.78      17.00   3.222   0.0105 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 524.9 on 9 degrees of freedom
Multiple R-squared:  0.5357,	Adjusted R-squared:  0.4841
F-statistic: 10.38 on 1 and 9 DF,  p-value: 0.01045
```

(a) Wie viele Neugeborene 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 des Geburtsgewichts wird durch das lineare Modell erklärt?

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

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

(g) Welches Geburtsgewicht prognostiziert das Modell für einen Fötus mit einem Bauchdurchmesser von 110 (in Millimetern)?

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

Zur Schätzung eines einfachen linearen Regressionsmodells

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

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

$$\sum_{i=1}^{25} y_i = 310.622; \quad \sum_{i=1}^{25} y_i^2 = 4578.739; \quad \sum_{i=1}^{25} x_i = 136.053;$$

$$\sum_{i=1}^{25} x_i^2 = 810.846; \quad \sum_{i=1}^{25} x_i \cdot y_i = 1810.876$$

(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 positiv 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 $x_0 = 7.5$ an.

5 Schließende Statistik WS 2016/17

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. Für die Erwartungswerte der Zufallsvariablen $X_1, \ldots, X_n$ und Y gelte $\mathrm{E}(X_1) = \ldots = \mathrm{E}(X_n) = \mathrm{E}(Y)$ sowie für die Varianzen $\mathrm{Var}(X_1) = \ldots = \mathrm{Var}(X_n) = \mathrm{Var}(Y)$. Dann ist $X_1, \ldots, X_n$ stets eine einfache Stichprobe zu Y. ☐ ☐

2. Ist eine Familie von Schätzfunktionen $T_n, n \in \mathbb{N}$, konsistent im quadratischen Mittel für einen Parameter $\lambda \in \mathbb{R}$, so gilt stets ☐ ☐
$$\lim_{n\to\infty} \mathrm{E}(T_n) = \lambda.$$

3. Ist $X_1, \ldots, X_n$ eine einfache Stichprobe zu einer mit bekannter Varianz und unbekanntem Erwatzungswert verteilten Zufallsvariablen Y, so erhält man bei einer Stichprobenziehung mit Wahrscheinlichkeit $1-\alpha$ eine Stichprobenrealisation, die zu einem realisierten Konfidenzintervall (zum Konfidenzniveau $1-\alpha$) für den Erwartungswert führt, welches den unbekannten Erwartungswert von Y enthält ($\alpha \in (0,1)$). ☐ ☐

4. Je kleiner der p-Wert zu einer realisierten Teststatistik ist, umso mehr spricht die realisierte Teststatistik für die Verletzung der Nullhypothese. ☐ ☐

5. Beim Test zum Vergleich von zwei Anteilswerten (als Spezialfall des 2-Stichproben-t-Tests) muss die Voraussetzung $\sigma_A^2 = \sigma_B^2$ nicht überprüft werden, da sie unter H_0 automatisch erfüllt ist. ☐ ☐

6. Wenn bei einem statistischen Test zum Signifikanzniveau α die Testentscheidung für H_0 ausfällt, dann gilt H_0 mindestens mit einer Wahrscheinlichkeit von $1-\alpha$. ☐ ☐

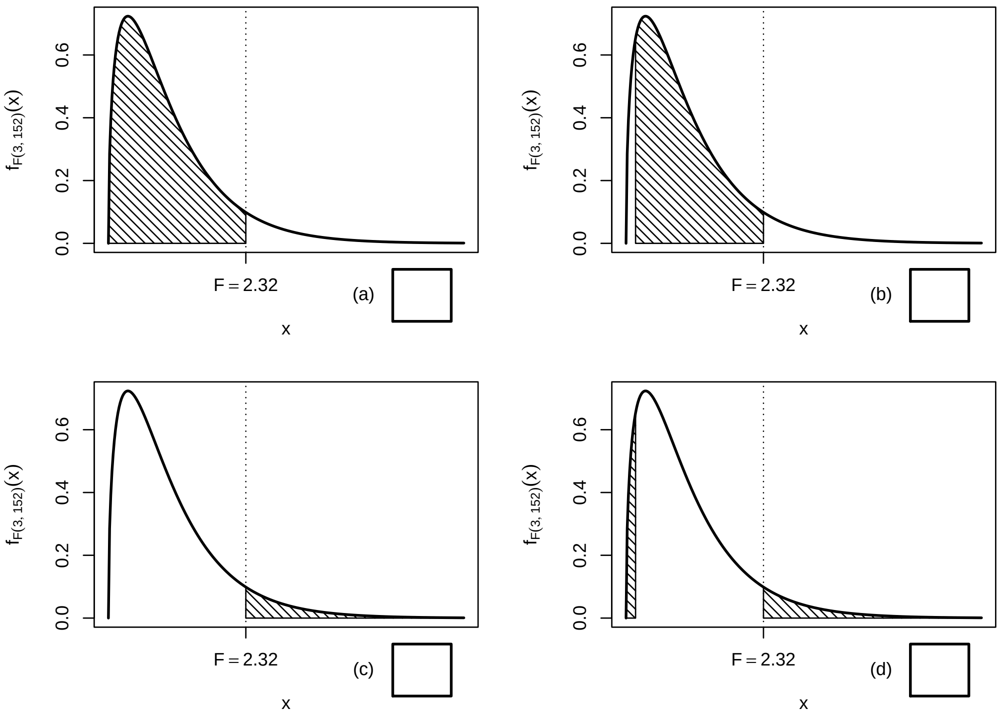
7. Sind die Voraussetzungen zur Anwendung der einfachen Varianzanalyse erfüllt und liegen als Stichprobenrealisation jeweils 25 Beobachtungen zu 6 Faktorstufen vor, so ist die Teststatistik bei Gültigkeit der Nullhypothese $F(5, 144)$-verteilt. ☐ ☐

8. Im einfachen linearen Regressionsmodell ☐ ☐
$$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2),$$
sind Prognoseintervalle für y_0 gegeben x_0 umso breiter, je weiter x_0 von $\overline{x}$ 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 mit $k = 4$ Faktorstufen und einem Gesamtstichprobenumfang von $n = 156$ erhält man die realisierte Teststatistik $F = 2.32$. 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.



2. Es sei $X_1, \ldots, X_{36}$ eine einfache Stichprobe vom Umfang 36 zu Y mit $Y \sim N(102, 3^2)$. Dann gilt für die Teststatistik $N = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$ des Gauß-Tests für den Mittelwert einer normalverteilten Zufallsvariablen mit bekannter Varianz zur Nullhypothese $H_0 : \mu = 100$:

(a) $N \sim N(0, 1)$ ☐

(b) $N \sim N(4, 1)$ ☐

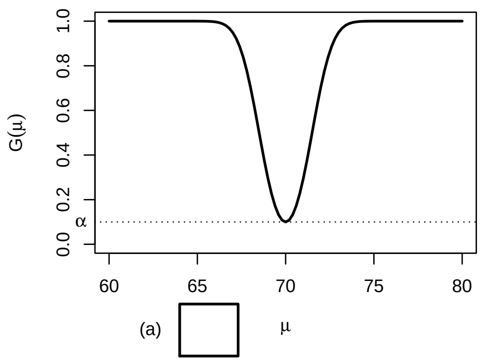
(c) $N \sim N(0, 3^2)$ ☐

(d) $N \sim N(4, 3^2)$ ☐

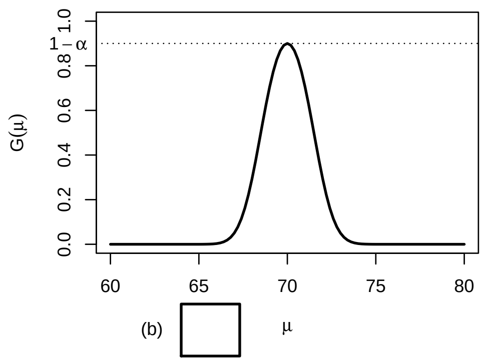
3. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{30}$ vom Umfang $n = 30$ zu einer $N(\mu, 5^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu = 70 \qquad \text{gegen} \qquad H_1 : \mu \neq 70$$

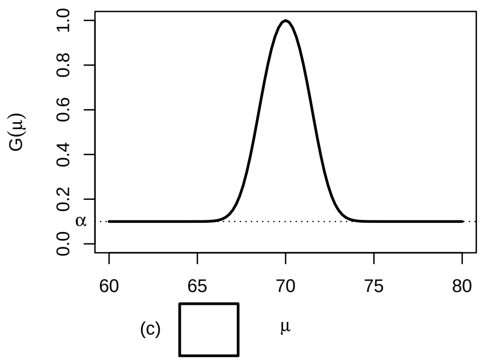
bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.
Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.



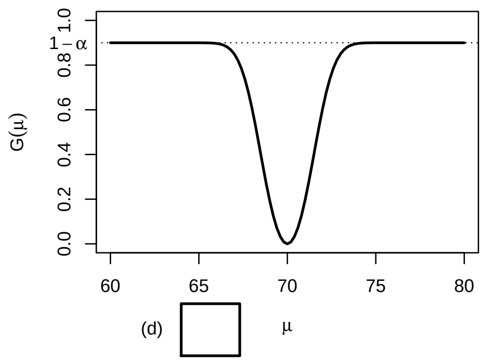
(a) ☐



(b) ☐



(c) ☐



(d) ☐

4. Als p-Wert zur realisierten Teststatistik eines rechtsseitigen Gauß-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz (mit $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$) erhält man $p = 0.0324$. Dann gilt für die p-Werte des linksseitigen Tests (mit $H_0 : \mu \geq \mu_0$ 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 linksseitigen Tests beträgt 0.0648, der p-Wert des zweiseitigen Tests 0.0162. ☐

   (b) Der p-Wert des linksseitigen Tests beträgt 0.0162, der p-Wert des zweiseitigen Tests 0.0648. ☐

   (c) Der p-Wert des linksseitigen Tests beträgt 0.9676, der p-Wert des zweiseitigen Tests 0.0162. ☐

   (d) Der p-Wert des linksseitigen Tests beträgt 0.9676, der p-Wert des zweiseitigen Tests 0.0648. ☐

Aufgabe 3 (8 Punkte)

Für $\sigma > 0$ sei die Zufallsvariable Y Rayleigh-verteilt mit Parameter σ. Es gilt dann $\mathrm{E}(Y) = \sqrt{\frac{\pi}{2}}\sigma$, $\mathrm{Var}(Y) = \frac{4-\pi}{2}\sigma^2$ sowie $\mathrm{E}(Y^4) = 8\sigma^4$. $X_1, \ldots, X_n$ sei für $n \in \mathbb{N}$ eine einfache Stichprobe vom Umfang n zu Y. Untersuchen Sie, ob die Schätzfunktionen

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

konsistent im quadratischen Mittel für σ^2 sind.

Aufgabe 4 (6 + 2 = 8 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $\lambda > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|\lambda) = \begin{cases} \lambda^2 \cdot (y+1) \cdot e^{-\lambda \cdot (y+1)} & \text{für } y > -1 \\ 0 & \text{sonst} \end{cases}$$

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

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

(b) Man kann zeigen, dass $\mathrm{E}(Y) = \frac{2}{\lambda} - 1$ gilt. Bestimmen Sie damit den Schätzer $\widehat{\lambda}_{MM}$ nach der Methode der Momente.

Hinweise:

- *Beachten Sie, dass Sie Teil (b) auch ohne die Bearbeitung von Teil (a) 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.*

Aufgabe 5 (3 + 7 + 4 = 14 Punkte)

Bei der Abfüllung von Fertigspachtel weiß der Hersteller aus Erfahrung, dass die verwendete Maschine eine Standardabweichung von $0.2[kg]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Maschine im Mittel eine andere Menge als die auf dem Produkt ausgezeichneten $10[kg]$ in die Eimer einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 9 Eimer entnommen, deren gemessene Füllmengen $x_1, \ldots, x_9$ als Realisation einer einfachen Stichprobe vom Umfang 9 zur annahmegemäß $N(\mu, 0.2^2[kg^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = 9.877[kg] \ .$$

(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 $\alpha = 0.10$ **(!)**, ob der Verdacht der Herstellerfirma bestätigt werden kann. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(c) 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 $10.12[kg]$ beträgt?

Aufgabe 6 (11 + 9 + 8 = 28 Punkte)

Um zu überprüfen, ob sich die Leistungsfähigkeit von NiMH-Akkus zweier verschiedener Marken unterscheidet, lässt ein Testinstitut die Ausdauer jeweils eines Akkusatzes in 10 unterschiedlichen Taschenlampenmodellen untersuchen. Es wurden dabei die folgenden Leuchtdauern (in Minuten) bis zur automatischen Abschaltung der Lampen festgestellt:

Taschenlampe i	1	2	3	4	5	6	7	8	9	10
Marke A x_i^A	232	270	249	256	194	244	235	236	257	254
Marke B x_i^B	216	241	254	223	226	235	210	242	227	225

(a) Überprüfen Sie unter der Annahme, dass die gemessenen Leuchtdauern (in Minuten) aus einer einfachen Stichprobe zur zweidimensional normalverteilten Grundgesamtheit (Y^A, Y^B) der Leuchtdauern mit Akkumarke A (Y^A) bzw. Akkumarke B (Y^B) stammen, zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass die Verwendung von Akkumarke A im Vergleich zu Akkumarke B durchschnittlich eine längere Leuchtdauer 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 Leuchtdauern zu den jeweiligen Taschenlampenmodellen verloren gegangen ist. Um die Situation zu retten, nehme man weiter an, dass mit $X_1^A, \ldots, X_{10}^A$ und $X_1^B, \ldots, X_{10}^B$ 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 Akkumarke A im Vergleich zu Akkumarke B eine längere Leuchtdauer ermöglicht. Verwenden Sie hierzu die Stichprobenmittelwerte $\overline{x^A} = 242.7$ bzw. $\overline{x^B} = 229.9$ sowie die Stichprobenvarianzen $s^2_{Y^A} = 431.79$ bzw. $s^2_{Y^B} = 173.43$. Fassen Sie das Ergebnis des Tests auch in einem Antwortsatz zusammen.

(c) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.10$, ob die in Teil (b) getroffene Annahme der Varianzgleichheit auf Grundlage der vorhandenen Stichprobeninformation verworfen werden muss. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (c) den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m,n)$*-Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

Aufgabe 7 (11 Punkte)

Mit einem Chi-Quadrat-Anpassungstest soll zum Signifikanzniveau $\alpha = 0.05$ überprüft werden, ob man bei einem beobachteten Stichprobenergebnis von der Realisation einer einfachen Stichprobe vom Umfang $n = 100$ zu einer Geom(0.45)-verteilten Zufallsvariablen ausgehen kann. Die Stichprobeninformation liege in Form der folgenden Häufigkeitsverteilung vor:

a_i	0	1	2	≥ 3
n_i	30	26	16	28

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

Hinweise:

- *Die geometrische Verteilung mit Parameter* $p = 0.45$ *hat den Träger* $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$ *und die Wahrscheinlichkeitsfunktion*

$$p_{\text{Geom}(0.45)} : \mathbb{N}_0 \to [0, 1]; p_{\text{Geom}(0.45)}(i) = (1 - 0.45)^i \cdot 0.45 \ .$$

- *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

Aufgabe 8 (1 + 1 + 1 + 1 + 1 = 5 Punkte)

Zur Erklärung des Umsatzes im Bauhauptgewerbe in Deutschland y_i (in Milliarden Euro) durch die Lohn- und Gehaltssumme im Bauhauptgewerbe in Deutschland 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 2005–2012 wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
    Min      1Q  Median      3Q     Max
-4.8207 -1.1130 -0.0814  1.8445  3.2378
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -29.558     21.821  -1.355  0.22434
x              6.038      1.149   5.256  0.00191 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.82 on 6 degrees of freedom
Multiple R-squared:  0.8216,       Adjusted R-squared:  0.7918
F-statistic: 27.63 on 1 and 6 DF,  p-value: 0.001908
```

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

(b) Welcher Anteil der Gesamtvarianz des Umsatzes im Bauhauptgewerbe in Deutschland wird durch das lineare Modell erklärt?

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

(d) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.001$, ob β_2 signifikant positiv ist.

(e) Welche Umsatzhöhe im Bauhauptgewerbe (in Milliarden Euro) prognostiziert das Modell für ein Jahr mit einer Lohn- und Gehaltssumme von 20 (in Milliarden Euro)?

Aufgabe 9 (6 + 2 + 2 + 3 + 5 = 18 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:

$$\sum_{i=1}^{25} y_i = 224.942; \quad \sum_{i=1}^{25} y_i^2 = 2625.587; \quad \sum_{i=1}^{25} x_i = 87.542;$$

$$\sum_{i=1}^{25} x_i^2 = 354.961; \quad \sum_{i=1}^{25} x_i \cdot y_i = 666.095$$

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

6 Schließende Statistik SS 2017

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 $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu einer normalverteilten Zufallsvariablen Y. Dann sind die Zufallsvariablen $X_1, \ldots, X_n$ stets unkorreliert.	☐	☐
2. Ist eine Schätzfunktion $\widehat{\theta}$ effizient in einer Klasse erwartungstreuer Schätzfunktionen für den Parameter $\theta \in \Theta$, so ist die Varianz jeder anderen Schätzfunktion aus dieser Klasse mindestens so groß wie die Varianz von $\widehat{\theta}$; dies gilt unabhängig davon, welches $\theta \in \Theta$ der wahre Parameter ist.	☐	☐
3. Zur Schätzung des Parameters $\theta \in \mathbb{R}$ seien für $n \in \mathbb{N}$ Schätzfunktionen T_n gegeben mit den Eigenschaften $\mathrm{E}(T_n) = \frac{\theta+1}{n}$ und $\mathrm{Var}(T_n) = \frac{3}{n}$ für alle $n \in \mathbb{N}$. Dann ist die Folge T_n von Schätzfunktionen für θ konsistent im quadratischen Mittel.	☐	☐
4. Die Nullhypothese $H_0 : \mu = \mu_0$ wird beim Gauß-Test auf den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz zum Signifikanzniveau α genau dann angenommen, wenn μ_0 im entsprechenden (symmetrischen) Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1-\alpha$ für μ enthalten ist.	☐	☐
5. Bei einem statistischen Hypothesentest ist die minimale Fehlerwahrscheinlichkeit 2. Art stets größer als die maximale Fehlerwahrscheinlichkeit 1. Art.	☐	☐
6. Liegt die Teststatistik T im kritischen Bereich eines Signifikanztests zum Signifikanzniveau α, so gilt für den p-Wert p zur Teststatistik T die Beziehung $p \geq \alpha$.	☐	☐
7. Mit einem Chi-Quadrat-Anpassungstest soll anhand einer einfachen Stichprobe vom Umfang $n = 200$ überprüft werden, ob die Grundgesamtheit normalverteilt ist. Nach Wahl einer geeigneten Klassierung aus 6 Klassen werden dazu zunächst die beiden unbekannten Parameter der Normalverteilung durch eine ML-Schätzung aus den klassierten Daten ermittelt. Damit ist zur Konstruktion des Ablehnungsbereiches die χ^2-Verteilung mit 4 Freiheitsgraden zu verwenden.	☐	☐
8. Bei der Anwendung der Kleinst-Quadrate-Methode wird die Summe der quadrierten vertikalen Abstände der Beobachtungspunkte zur Regressionsgeraden minimiert.	☐	☐

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. Es sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y mit $Y \sim N(\mu, \sigma^2)$. Dann gilt für die Verteilung von $\sum_{i=1}^{n} X_i$:

   (a) $\sum_{i=1}^{n} X_i \sim N\left(\mu, \sigma^2\right)$ ☐

   (b) $\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ☐

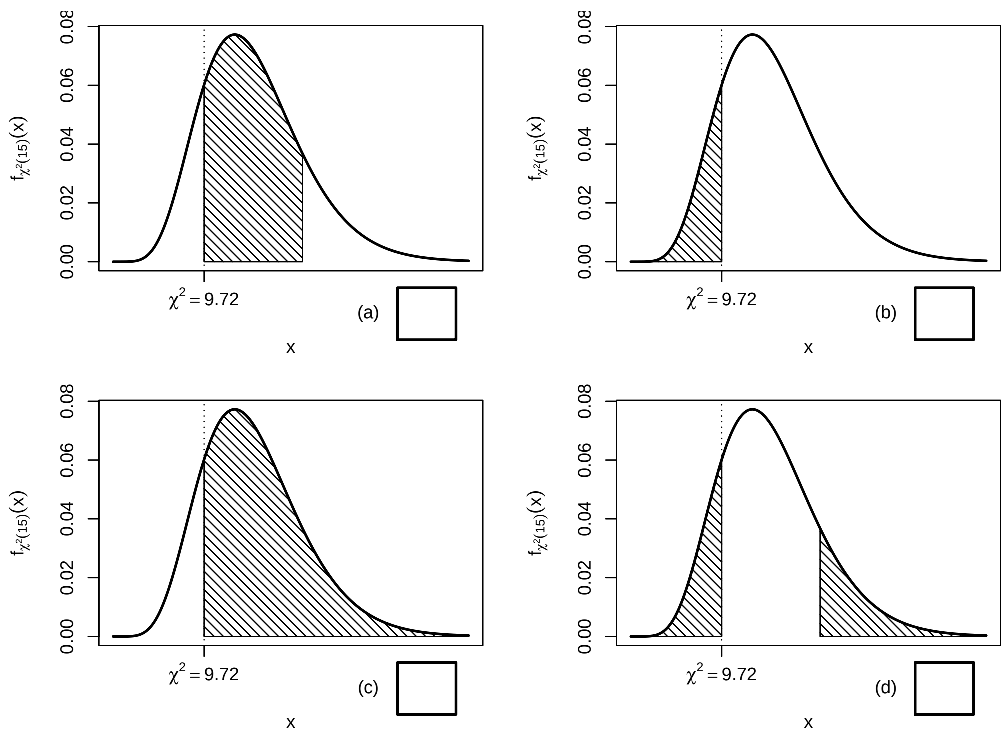
   (c) $\sum_{i=1}^{n} X_i \sim N\left(n\mu, \sigma^2\right)$ ☐

   (d) $\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$ ☐

2. Sei $X_1, \ldots, X_{16}$ 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

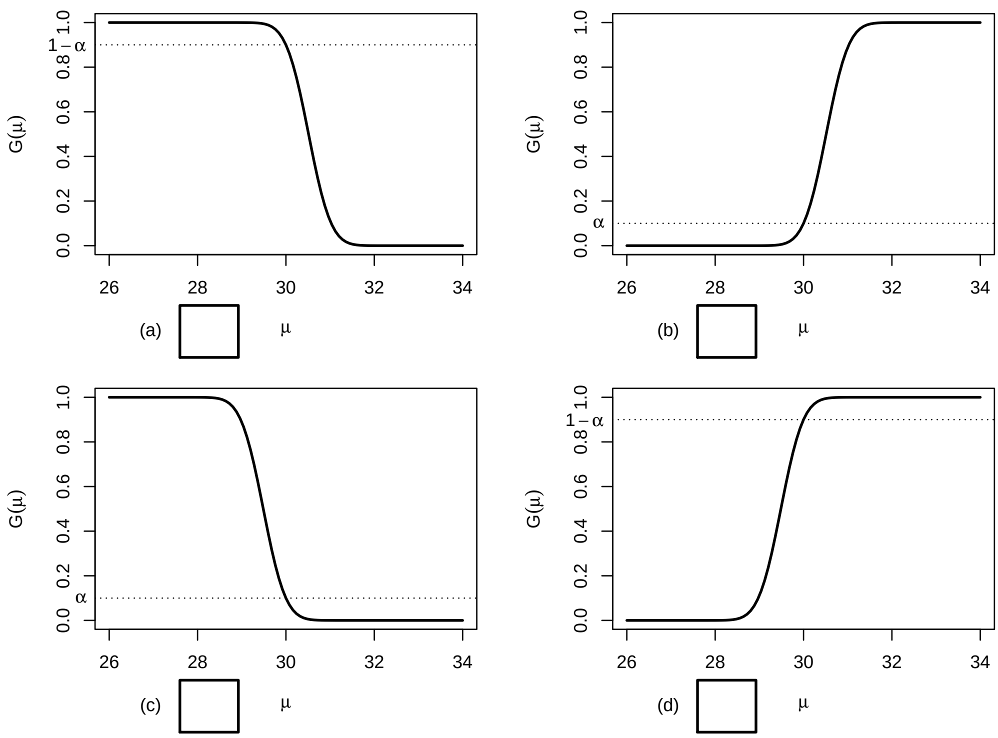
$$H_0 : \sigma^2 \geq \sigma_0^2 = 16 \qquad \text{gegen} \qquad H_1 : \sigma^2 < \sigma_0^2 = 16$$

   mit einem Chi-Quadrat-Test getestet werden. Als realisierte Teststatistik erhält man $\chi^2 = 9.72$. 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. Als p-Wert zur realisierten Teststatistik eines zweiseitigen Gauß-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz (mit $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$) erhält man $p = 0.04759$. 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) Auf Grundlage der vorhandenen Informationen ist noch unklar, ob bei keinem, genau einem oder beiden einseitigen Tests H_0 abgelehnt wird. ☐
   (b) Bei beiden einseitigen Tests wird H_0 abgelehnt. ☐
   (c) Bei keinem der beiden einseitigen Tests wird H_0 abgelehnt. ☐
   (d) Bei genau einem der beiden einseitigen Tests wird H_0 abgelehnt. Bei welchem dies der Fall ist, hängt vom Vorzeichen der Teststatistik ab. ☐

4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{25}$ vom Umfang $n = 25$ zu einer $N(\mu, 2^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen
$$H_0 : \mu \leq 30 \qquad \text{gegen} \qquad H_1 : \mu > 30$$
bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.
Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.



Aufgabe 3 (5 Punkte)

Für $0 < p < 1$ sei $Y \sim \text{Geom}(p)$, es gilt also insbesondere $\text{E}(Y) = \frac{1-p}{p}$ sowie $\text{Var}(Y) = \frac{1-p}{p^2}$. $X_1, \ldots, X_n$ sei für $n \in \mathbb{N}$ eine einfache Stichprobe vom Umfang n zu Y. Untersuchen Sie, ob die Schätzfunktionen
$$T_n(X_1, \ldots, X_n) := \frac{1}{2n} \sum_{i=1}^{n} \left(X_i^2 + X_i\right)$$

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

Aufgabe 4 (6 + 2 = 8 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $b > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|b) = \begin{cases} \frac{b^3}{2} \cdot (y-1)^2 \cdot e^{-b \cdot (y-1)} & \text{für } y > 1 \\ 0 & \text{sonst} \end{cases}$$

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) Man kann zeigen, dass $\mathrm{E}(Y) = \frac{3}{b} + 1$ gilt. Bestimmen Sie damit den Schätzer $\widehat{b}_{MM}$ nach der Methode der Momente.

Hinweise:

- *Beachten Sie, dass Sie Teil (b) auch ohne die Bearbeitung von Teil (a) 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.*

Aufgabe 5 (3 + 7 + 4 = 14 Punkte)

Bei der Abfüllung von Estrichbeton weiß der Hersteller aus Erfahrung, dass die verwendete Maschine eine Standardabweichung von $0.2[kg]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Maschine im Mittel mehr als die auf dem Produkt ausgezeichneten $40[kg]$ in die Säcke einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 16 Säcke entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{16}$ als Realisation einer einfachen Stichprobe vom Umfang 16 zur annahmegemäß $N(\mu, 0.2^2[kg^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 40.091[kg] \ .$$

(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) 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 $40.1[kg]$ beträgt?

Aufgabe 6 (9 + 7 = 16 Punkte)

Ein Hersteller von Metallwaren produziert unter anderem Flachverbinder mit einer Soll-Länge von 15 [cm]. Es soll angenommen werden, dass die Länge der hergestellten Flachverbinder gemäß einer Normalverteilung mit unbekanntem Erwartungswert und unbekannter Varianz schwankt. Zur Überprüfung der Qualität der produzierten Flachverbinder werden 10 Flachverbinder aus dem laufenden Produktionsprozess entnommen. Die gemessenen Längen

$$14.85, 14.94, 14.93, 14.87, 15.03, 14.93, 15.05, 15.12, 14.98, 14.99$$

seien als Realisation einer einfachen Stichprobe zur normalverteilten Grundgesamtheit aufzufassen. Aus dieser Realisation wurde bereits $s^2 = 0.006832$ berechnet.

(a) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die tatsächliche mittlere Länge der produzierten Flachverbinder im Vergleich zur angegebenen Soll-Länge von 15 [cm] zu klein ist. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die Varianz der Länge der produzierten Flachverbinder im Vergleich zur vom Hersteller angegebenen Toleranz $\sigma_0^2 = 0.0025$ zu groß ist. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) 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
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

Aufgabe 7 (10 Punkte)

Zwei unterschiedlichen Gruppen mit 56 (Gruppe A) bzw. 66 (Gruppe B) Heuschnupfenpatienten wird jeweils ein spezielles Antihistaminikum verabreicht. Nach einer festgelegten Zeit werden dann alle Heuschnupfenpatienten gefragt, ob durch das verabreichte Medikament eine Minderung der Beschwerden eingetreten ist. In der Gruppe der Heuschnupfenpatienten, denen Antihistaminikum A verabreicht wurde, beantworten 34 Personen diese Frage positiv, in der zu Antihistaminikum B gehörigen Gruppe 49 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 Antihistaminika unterscheidet (bezogen auf die Erfolgswahrscheinlichkeit für eine Minderung der Beschwerden). Formulieren Sie das Ergebnis auch in Form eines Antwortsatzes.

Aufgabe 8 (13 Punkte)

Um die Leistungsfähigkeit von 4 Schulklassen einer Klassenstufe zu vergleichen, soll anhand der Ergebnisse einer Vergleichsarbeit untersucht werden, ob die Verteilung der von den Schülern erreichten Punktzahlen abhängig davon ist, welcher der 4 Klassen sie angehören. Zu den verschiedenen Schulklassen wurden die folgenden (fiktiven) Daten zu den erreichten Punktzahlen erhoben:

j (Klasse)	n_j	$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{j,i}$	$\sum_{i=1}^{n_j} x_{j,i}^2$
1	22	15.995	6034.40
2	21	20.424	9128.66
3	28	15.601	7401.69
4	26	17.208	8050.87

Für die Durchführung der einfachen Varianzanalyse wurde hieraus bereits die Größe $SW = 1713.254$ berechnet.
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 $X_{j,i}(1 \leq j \leq 4, 1 \leq i \leq n_j)$ sind, ob die Zugehörigkeit zu den unterschiedlichen Schulklassen einen Einfluss auf die (mittlere) erreichte Punktzahl hat. Fassen Sie das Ergebnis der Varianzanalyse auch in einem Antwortsatz zusammen.

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

$n \backslash m$	1	2	3	4	5	90	91	92	93	94
1	161.448	199.500	215.707	224.583	230.162	252.900	252.915	252.931	252.945	252.960
2	18.513	19.000	19.164	19.247	19.296	19.485	19.485	19.485	19.485	19.485
3	10.128	9.552	9.277	9.117	9.013	8.557	8.557	8.556	8.556	8.556
4	7.709	6.944	6.591	6.388	6.256	5.668	5.668	5.667	5.667	5.666
5	6.608	5.786	5.409	5.192	5.050	4.409	4.409	4.409	4.408	4.408
90	3.947	3.098	2.706	2.473	2.316	1.417	1.416	1.415	1.414	1.413
91	3.946	3.097	2.705	2.472	2.315	1.415	1.414	1.413	1.412	1.411
92	3.945	3.095	2.704	2.471	2.313	1.414	1.413	1.412	1.411	1.410
93	3.943	3.094	2.703	2.470	2.312	1.412	1.411	1.410	1.409	1.408
94	3.942	3.093	2.701	2.469	2.311	1.411	1.410	1.409	1.407	1.406

Aufgabe 9 ($1 + 1 + 1 + 1 + 1 + 1 + 1 = 7$ Punkte)

Zur Erklärung des Blutdrucks y_i durch das Verhältnis von tatsächlichem Gewicht zum Idealgewicht 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 den Daten einer US-amerikanischen Studie mit ausschließlich weiblichen Probanden wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
    Min      1Q  Median      3Q     Max
-21.202 -13.419  -2.503   5.834  70.957

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    72.74      19.62   3.708 0.000845 ***
x              37.17      13.59   2.734 0.010383 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.54 on 30 degrees of freedom
Multiple R-squared:  0.1995,	Adjusted R-squared:  0.1728
F-statistic: 7.477 on 1 and 30 DF,  p-value: 0.01038
```

(a) Wie viele weibliche Testpersonen 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 des Blutdrucks wird durch das lineare Modell erklärt?

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

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

(g) Welchen Blutdruck prognostiziert das Modell für eine weibliche Person mit einem Verhältnis zwischen tatsächlichem Gewicht und Idealgewicht von 1.1?

Aufgabe 10 (6 + 2 + 3 + 3 + 5 = 19 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 $n = 30$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{30} y_i = 119.349; \quad \sum_{i=1}^{30} y_i^2 = 850.131; \quad \sum_{i=1}^{30} x_i = 180.079;$$

$$\sum_{i=1}^{30} x_i^2 = 1165.397; \quad \sum_{i=1}^{30} x_i \cdot y_i = 572.096$$

(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 symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_2 an.

(e) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für y_0 gegeben $x_0 = 4$ an.

7 Schließende Statistik WS 2017/18

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 $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y. Dann sind $X_1, \ldots, X_n$ stochastisch unabhängig und die Verteilung aller X_i stimmt mit der Verteilung von Y überein.	☐	☐
2. Gilt für eine Familie $\widehat{\theta}_n, n \in \mathbb{N}$, von Schätzfunktionen für einen Parameter $\theta \in \Theta \subseteq \mathbb{R}$ unabhängig von θ sowohl $\mathrm{E}(\widehat{\theta}_n) = \theta$ für jedes $n \in \mathbb{N}$ als auch $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) = 0$, dann ist die Familie $\widehat{\theta}_n$ konsistent im quadratischen Mittel für θ.	☐	☐
3. Konfidenzintervalle für den Erwartungswert bei unbekannter Varianz sind umso breiter, je größer die geschätzte (Stichproben-)Varianz ist.	☐	☐
4. Statistische Hypothesentests sind typischerweise so konstruiert, dass die Teststatistik bei Gültigkeit der Nullhypothese mit einer größeren Wahrscheinlichkeit im kritischen Bereich liegt als bei Verletzung der Nullhypothese.	☐	☐
5. Das Verkleinern des Signifikanzniveaus führt bei Anwendung des zweiseitigen t-Tests für den Erwartungswert bei unbekannter Varianz stets zu einer Verkleinerung des kritischen Bereichs.	☐	☐
6. Mit Hilfe von Gütefunktionswerten kann man in Abhängigkeit der wahren Verteilungsparameter sowohl die Wahrscheinlichkeit für einen Fehler als auch die Wahrscheinlichkeit für eine korrekte Entscheidung leicht berechnen. Für feste Verteilungsparameter addieren sich diese beiden Wahrscheinlichkeiten stets zu 1.	☐	☐
7. Mit der einfachen Varianzanalyse können mehrere mit übereinstimmender Varianz normalverteilte Grundgesamtheiten daraufhin untersucht werden, ob ihre Erwartungswerte (ebenfalls) übereinstimmen.	☐	☐
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 die Kleinst-Quadrate-Schätzfunktionen $\widehat{\beta}_1$ und $\widehat{\beta}_2$ stets stochastisch unabhängig.	☐	☐

Aufgabe 2 (12 Punkte)

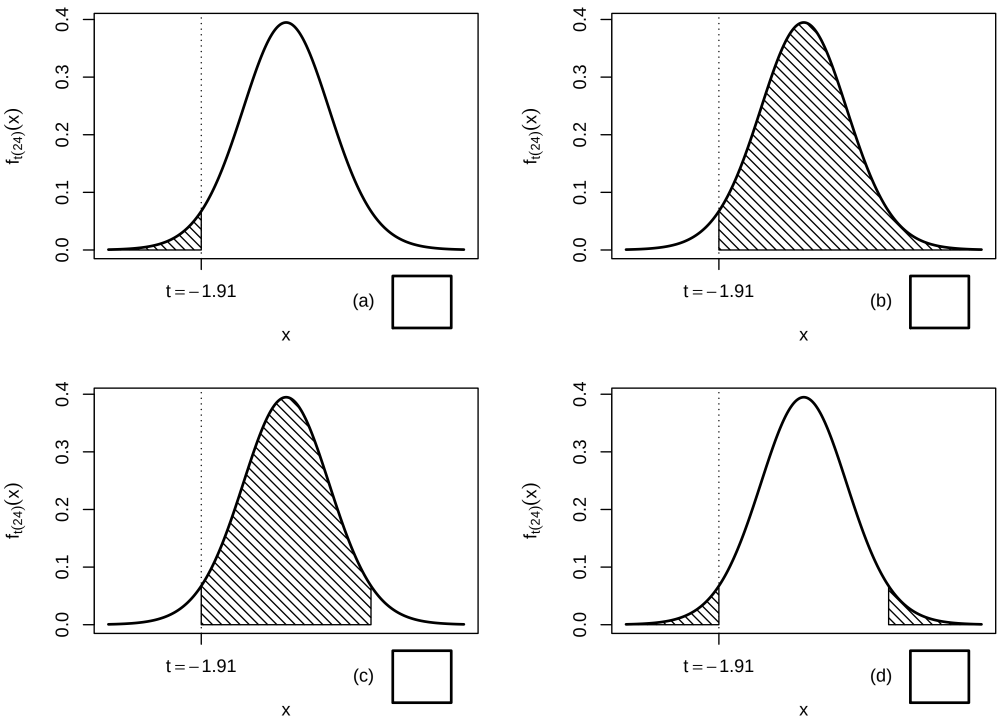
Markieren Sie jeweils die korrekte Antwort mit einem Kreuz im betreffenden Kästchen. Es ist jeweils genau ein Kreuz korrekt.

Richtige Antworten geben 3 Punkte, falsche Antworten und nicht bearbeitete Aufgabenteile 0 Punkte (Aufgabenteile mit mehr als einem Kreuz zählen als nicht bearbeitet!).

1. 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

$$H_0 : \mu = \mu_0 = 25 \qquad \text{gegen} \qquad H_1 : \mu \neq \mu_0 = 25$$

mit einem t-Test getestet werden. Als realisierte Teststatistik erhält man $t = -1.91$. 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.

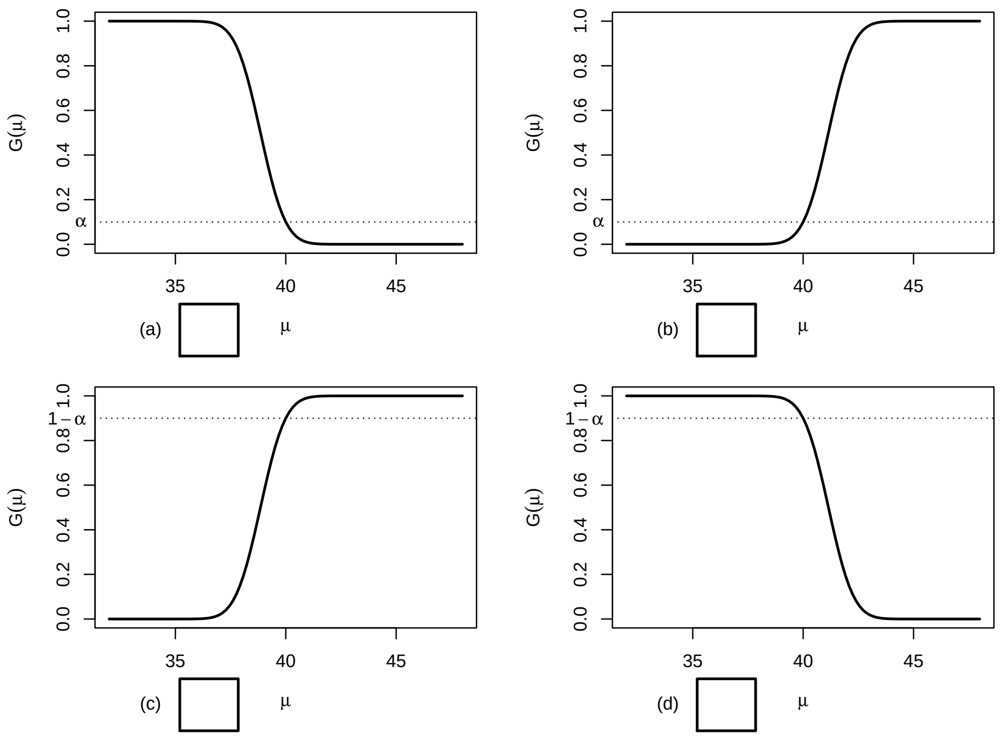


2. Auf Grundlage zweier unabhängiger einfacher Stichproben $X_1^A, \ldots, X_{n_A}^A$ vom Umfang n_A zu Y^A und $X_1^B, \ldots, X_{n_B}^B$ vom Umfang n_B zu Y^B soll unter der Annahme, dass Y^A und Y^B jeweils normalverteilt sind mit unbekannter, aber übereinstimmender Varianz, mit einem Signifikanztest überprüft werden, ob $\mathrm{E}(Y^A) < \mathrm{E}(Y^B)$ gilt. Zur Untersuchung dieser Fragestellung sind die folgenden aus der Vorlesung bekannten Verfahren geeignet:
   (a) Nur die Varianzanalyse ☐
   (b) Nur der 2-Stichproben-t-Test für den Mittelwert ☐
   (c) Nur der t-Differenzentest für verbundene Stichproben ☐
   (d) Die Varianzanalyse und (äquivalent) der 2-Stichproben-t-Test für den Mittelwert ☐

3. Auf der Grundlage einer einfachen Stichprobe $X_1, \dots, X_{20}$ vom Umfang $n = 20$ zu einer $N(\mu, 4^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu \geq 40 \qquad \text{gegen} \qquad H_1 : \mu < 40$$

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 rechtsseitigen Gauß-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz (mit $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$) wird die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ abgelehnt. Dann gilt für die Testentscheidungen des linksseitigen (mit $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$) und zweiseitigen (mit $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$) Tests (bei unverändertem Signifikanzniveau $\alpha = 0.05$):

   (a) Der linksseitige Test lehnt H_0 nicht ab, der zweiseitige Test lehnt H_0 ab. ☐

   (b) Der linksseitige Test lehnt H_0 ab, der zweiseitige Test lehnt H_0 nicht ab. ☐

   (c) Der linksseitige Test lehnt H_0 nicht ab, die Entscheidung des zweiseitigen Tests kann für oder gegen H_0 ausfallen. ☐

   (d) Der linksseitige Test lehnt H_0 ab, die Entscheidung des zweiseitigen Tests kann für oder gegen H_0 ausfallen. ☐

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

Zu $Y \sim N(\mu, \sigma^2)$ liegen die unabhängigen einfachen Stichproben $X_1^A, \dots, X_{30}^A$ vom Umfang 30 und $X_1^B, \dots, X_{10}^B$ vom Umfang 10 vor. Mit $\overline{X^A} := \frac{1}{30}\sum_{i=1}^{30} X_i^A$ und $\overline{X^B} := \frac{1}{10}\sum_{i=1}^{10} X_i^B$ werden die Schätzfunktionen

- $\widehat{\mu}_1 = \frac{1}{4} \cdot \overline{X^A} + \frac{1}{4} \cdot \overline{X^B}$,

- $\widehat{\mu}_2 = \frac{1}{4} \cdot \overline{X^A} + \frac{3}{4} \cdot \overline{X^B}$ und
- $\widehat{\mu}_3 = \frac{3}{4} \cdot \overline{X^A} + \frac{1}{4} \cdot \overline{X^B}$

zur Schätzung von μ betrachtet.

(a) Wie sind $\overline{X^A}$ und $\overline{X^B}$ verteilt?

(b) Welche der Schätzfunktionen $\widehat{\mu}_1$, $\widehat{\mu}_2$ und $\widehat{\mu}_3$ sind erwartungstreu für μ? Begründen Sie Ihre Antwort.

(c) Berechnen Sie zu den für μ erwartungstreuen Schätzfunktionen die zugehörige Varianz. Welche dieser Schätzfunktionen würden Sie am ehesten zur Schätzung von μ einsetzen? Begründen Sie Ihre Antwort.

Aufgabe 4 (6 + 3 + 1 = 10 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $a > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|a) = \begin{cases} \frac{3 \cdot a^3}{y^4} & \text{für } y \geq a \\ 0 & \text{sonst} \end{cases}$$

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 $E(Y) = \frac{3}{2} \cdot a$ 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.*

Aufgabe 5 (3 + 7 + 2 + 3 = 15 Punkte)

Bei der Abfüllung von Nasenspray weiß der Hersteller aus Erfahrung, dass die verwendete Maschine eine Standardabweichung von $0.15[ml]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Maschine im Mittel weniger als die auf dem Produkt ausgezeichneten $10[ml]$ in die Sprühflaschen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 9 Sprühflaschen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_9$ als Realisation einer einfachen Stichprobe vom Umfang 9 zur annahmegemäß $N(\mu, 0.15^2[ml^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{9} \sum_{i=1}^{9} x_i = 9.875[ml] \ .$$

(a) Geben Sie auf Basis der Stichprobenrealisation ein zweiseitiges Konfidenzintervall für die mittlere Abfüllmenge zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.99(!)$ 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 Gegenhypothese zu erhalten, falls die tatsächliche mittlere Abfüllmenge $10.05[ml]$ beträgt?

Aufgabe 6 (10 + 8 = 18 Punkte)

Es werde angenommen, dass die (in mAh bei festem Entladestrom gemessene) Kapazität Y^A des aktuell von einem Taschenlampenhersteller verwendeten Lithium-Ionen-Akkutyps normalverteilt sei mit unbekanntem Erwartungswert μ_A und unbekannter Varianz σ_A^2. Der Taschenlampenhersteller erwägt, seine Taschenlampen zukünftig mit einem alternativen Lithium-Ionen-Akkutyp auszuliefern, dessen Kapazität Y^B ebenfalls als normalverteilt (mit unbekanntem Erwartungswert μ_B und unbekannter Varianz σ_B^2) angenommen werden kann. Es soll überprüft werden, ob der alternative Akkutyp im Mittel eine höhere Kapazität als der aktuell verwendete Typ besitzt.
Aus einer Kapazitätsmessung mit $n_A = 8$ Exemplaren des aktuell verwendeten Akkutyps und $n_B = 10$ Exemplaren des alternativen Akkutyps erhielt man Realisationen jeweils voneinander unabhängiger einfacher Stichproben $X_1^A, \ldots, X_8^A$ zu Y^A sowie $X_1^B, \ldots, X_{10}^B$ zu Y^B und hieraus die zugehörigen Mittelwerte $\overline{x^A} = 2984$ bzw. $\overline{x^B} = 3134$ sowie die Stichproben**standardabweichungen** $s_{Y^A} = 117$ bzw. $s_{Y^B} = 106$.

(a) Testen Sie unter der Annahme $\sigma_A^2 = \sigma_B^2$ zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass der alternative Akkutyp im Mittel eine höhere Kapazität als der aktuell verwendete Typ besitzt. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.10$, ob die in Teil (a) getroffene Annahme der Varianzgleichheit auf Grundlage der vorhandenen Stichprobeninformation verworfen werden muss. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) den folgenden Tabellenausschnitt mit 0.95-*Quantilen von* $F(m, n)$-*Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

Aufgabe 7 (10 + 3 + 2 = 15 Punkte)

In einer bestimmten Multiple-Choice-Aufgabe ist genau eine von vier Antwortmöglichkeiten, die mit „A“, „B“, „C“ bzw. „D“ bezeichnet sind, korrekt. Bei der Korrektur der Klausur stellt der Dozent fest, dass die von den Teilnehmern der Klausur abgegebenen Antworten wie folgt verteilt sind:

Antwort	A	B	C	D
Anteil in Prozent	20	30	18	32

Der Dozent fragt sich, ob man bei dieser Verteilung der abgegebenen Antworten davon ausgehen kann, dass sich die Teilnehmer der Klausur rein zufällig (und voneinander unabhängig) für eine der vier Antworten entschieden haben.

(a) Gehen Sie zunächst davon aus, dass 300 Studierende an der Klausur teilgenommen (und die Aufgabe bearbeitet) haben. Überprüfen Sie unter der Annahme, dass sich die abgegebenen Antworten als einfache Stichprobe auffassen lassen, mit einem geeigneten statistischen Test, ob sich die Prüflinge rein zufällig für eine der vier Antworten entschieden haben oder nicht (Signifikanzniveau $\alpha = 0.01$ **(!)**).

(b) Würde sich das Ergebnis des Tests aus Teil (a) ändern, wenn nur 150 Prüflinge an der Klausur teilgenommen (und die Aufgabe bearbeitet) hätten? Begründen Sie Ihre Antwort! (*Beachten Sie, dass Sie den Test nicht komplett neu durchführen müssen und insbesondere die realisierte Teststatistik relativ leicht aus dem Ergebnis des vorherigen Aufgabenteils gewinnen können!*)

(c) Ab welcher Anzahl von Klausurteilnehmern würde eine Durchführung des bereits in den Teilen (a) und (b) verwendeten Tests dazu führen, davon auszugehen, dass sich die Prüflinge *nicht* rein zufällig für eine der vier Antwortmöglichkeiten entschieden haben?

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

Aufgabe 8 (1 + 1 + 1 + 1 + 1 = 5 Punkte)

Zur Erklärung des Erdgaspreises je Mio. BTU y_i (in US-Dollar) durch den Erdölpreis je Barrel x_i (in US-Dollar) unterstellt man die Gültigkeit eines Zusammenhangs im Sinne des folgenden linearen Modells:

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

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 
-0.2309 -0.7273  0.1330  0.7239 -0.1521  1.1144 -0.8609 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.58394    1.02301   1.548 0.182227    
x            0.08710    0.01129   7.718 0.000583 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7912 on 5 degrees of freedom
Multiple R-squared:  0.9226,	Adjusted R-squared:  0.9071 
F-statistic: 59.56 on 1 and 5 DF,  p-value: 0.0005831
```

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

(d) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.001$, ob β_2 signifikant von Null verschieden ist.

(e) Welchen Erdgaspreis je Mio. BTU (in US-Dollar) prognostiziert das Modell für ein Jahr mit einem Erdölpreis je Barrel von 70 (in US-Dollar)?

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 $n = 20$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{20} y_i = 339.593; \quad \sum_{i=1}^{20} y_i^2 = 5903.002; \quad \sum_{i=1}^{20} x_i = 93.608;$$

$$\sum_{i=1}^{20} x_i^2 = 481.297; \quad \sum_{i=1}^{20} x_i \cdot y_i = 1632.131$$

(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 positiv ist. Fassen Sie das Ergebnis auch in einem Antwortsatz zusammen.

(e) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.90$ für y_0 gegeben $x_0 = 5$ an.

8 Schließende Statistik SS 2018

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!).

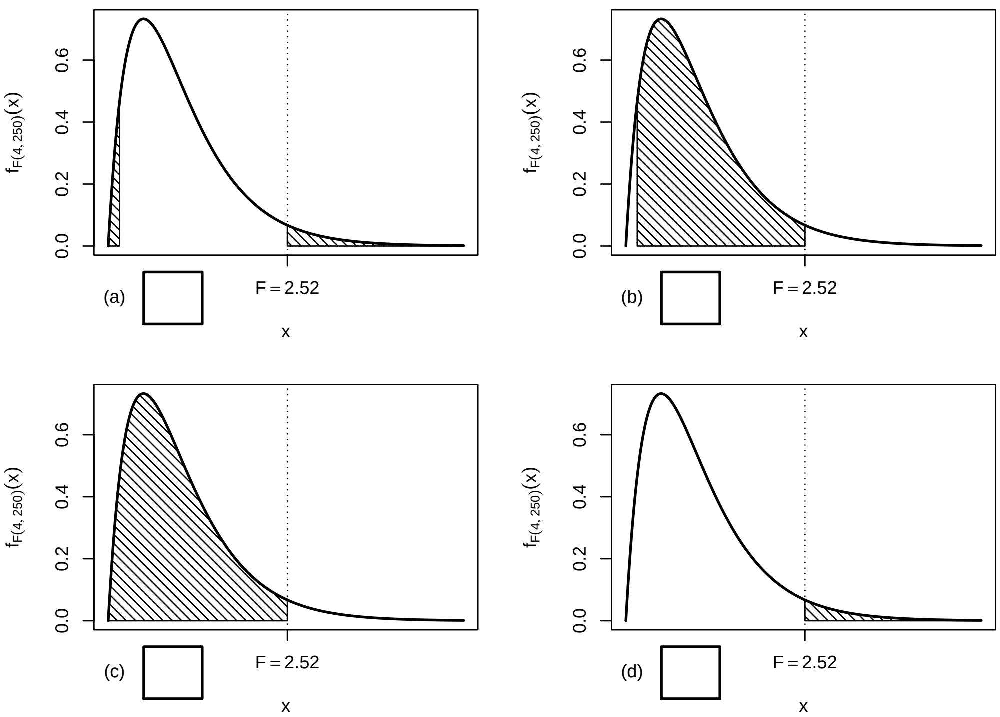
	wahr	falsch
1. Sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y. Dann sind $X_1, \ldots, X_n$ stets normalverteilt.	☐	☐
2. Sind für $n \in \mathbb{N}$ Schätzfunktionen T_n gegeben, für die $\mathrm{E}(T_n) = \lambda$ sowie $\lim_{n\to\infty} \mathrm{Var}(T_n) = 0$ gilt, dann ist die 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 $x_1, \ldots, x_n$ 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 $y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ 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!).

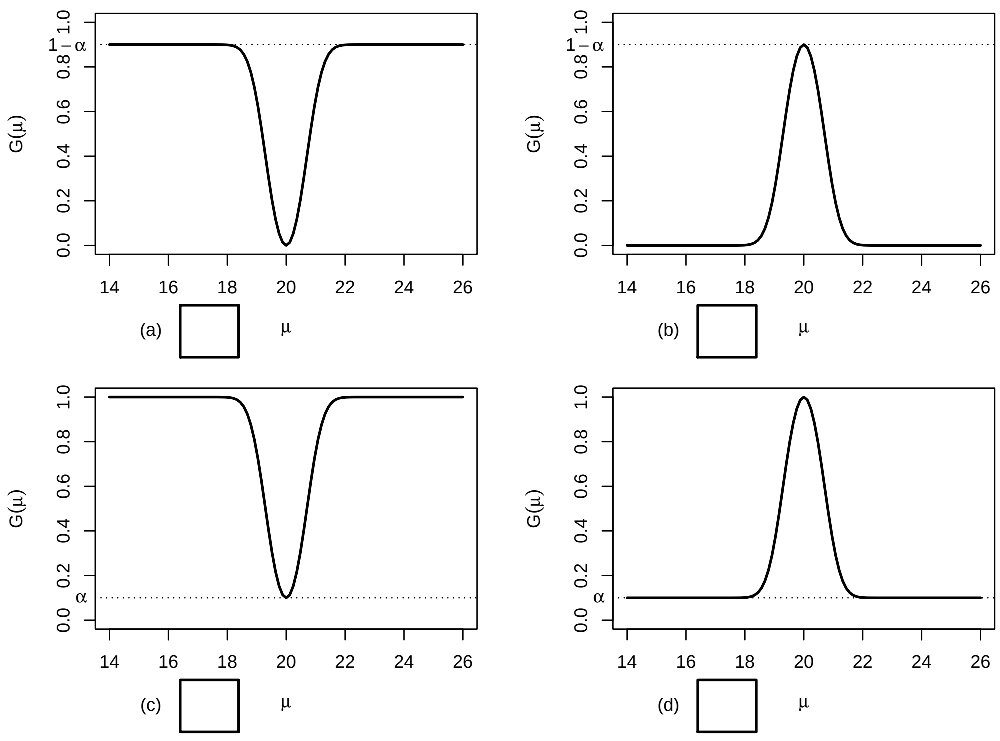
1. Auf Grundlage einer einfachen Stichprobe $X_1, \ldots, X_n$ vom Umfang n zu einer Zufallsvariablen Y, von der man lediglich weiß, dass sie normalverteilt ist, soll mit einem Signifikanztest überprüft werden, ob $\text{Var}(Y) \neq 3^2$ gilt. Zur Untersuchung dieser Fragestellung ist geeignet:
   (a) 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 $\mathrm{E}(Y) = \dfrac{2}{p}$ sowie $\mathrm{Var}(Y) = \dfrac{2}{p^2}$. 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

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

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

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

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 $X_1, \dots, X_n$ 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.*

Aufgabe 5 (7 + 2 + 4 + 4 = 17 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 $x_1, \dots, x_{16}$ 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?

(d) Wie groß muss der Stichprobenumfang bei der oben beschriebenen Qualitätskontrolle mindestens gewählt werden, wenn eine tatsächliche Überschreitung des Erwartungswerts der Länge um $1[cm]$ mit einer Wahrscheinlichkeit von mindestens 99% erkannt werden soll?

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.

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

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:

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

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)?

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

Zur Schätzung eines einfachen linearen Regressionsmodells

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

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.

9 Schließende Statistik WS 2018/19

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 Y, so ist der Stichprobenmittelwert $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ effizient in der Klasse aller linearen erwartungstreuen Schätzfunktionen für den Erwartungswert von Y.	☐	☐
2. Ist eine Familie $\widehat{\theta}_n, n \in \mathbb{N}$, von Schätzfunktionen konsistent im quadratischen Mittel für einen Parameter θ, so gilt für jedes $n \in \mathbb{N}$ stets $\mathrm{E}(\widehat{\theta}_n) = \theta$.	☐	☐
3. Die Breite von (approximativen) Konfidenzintervallen für den Parameter p einer alternativverteilten Zufallsvariablen ist besonders klein, wenn der Anteil der Erfolge $\widehat{p}$ in der Stichprobenrealisation in der Nähe von 50% liegt.	☐	☐
4. Lehnt ein zweiseitiger t-Test für den Erwartungswert einer normalverteilten Zufallsvariable bei unbekannter Varianz die Nullhypothese zum Signifikanzniveau $\alpha < 0.50$ ab, so wird auch stets genau einer der zugehörigen einseitigen t-Tests (bei unverändertem Signifikanzniveau) die Nullhypothese ablehnen.	☐	☐
5. Beim zweiseitigen Gauß-Test für den Erwartungswert bei bekannter Varianz (mit $H_0 : \mu = \mu_0$) ist der p-Wert umso höher, je größer der Abstand $\lvert\overline{X} - \mu_0\rvert$ ausfällt.	☐	☐
6. Mit einem Chi-Quadrat-Anpassungstest soll anhand einer einfachen Stichprobe vom Umfang $n = 200$ überprüft werden, ob die Grundgesamtheit exponentialverteilt ist. Nach Wahl einer geeigneten Klassierung aus 7 Klassen wird dazu zunächst der unbekannte Parameter der Exponentialverteilung durch eine ML-Schätzung aus den klassierten Daten ermittelt. Damit ist zur Konstruktion des Ablehnungsbereiches die χ^2-Verteilung mit 4 Freiheitsgraden zu verwenden.	☐	☐
7. Mit der einfachen Varianzanalyse kann untersucht werden, welche Ausprägung eines Faktors (Faktorstufe) zum höchsten Erwartungswert führt.	☐	☐
8. Im einfachen linearen Regressionsmodell $$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad i = 1, \dots, n,$$ gilt für die Residuen $\widehat{u}_i = y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_i)$ stets $\sum_{i=1}^{n} \widehat{u}_i = 0$.	☐	☐

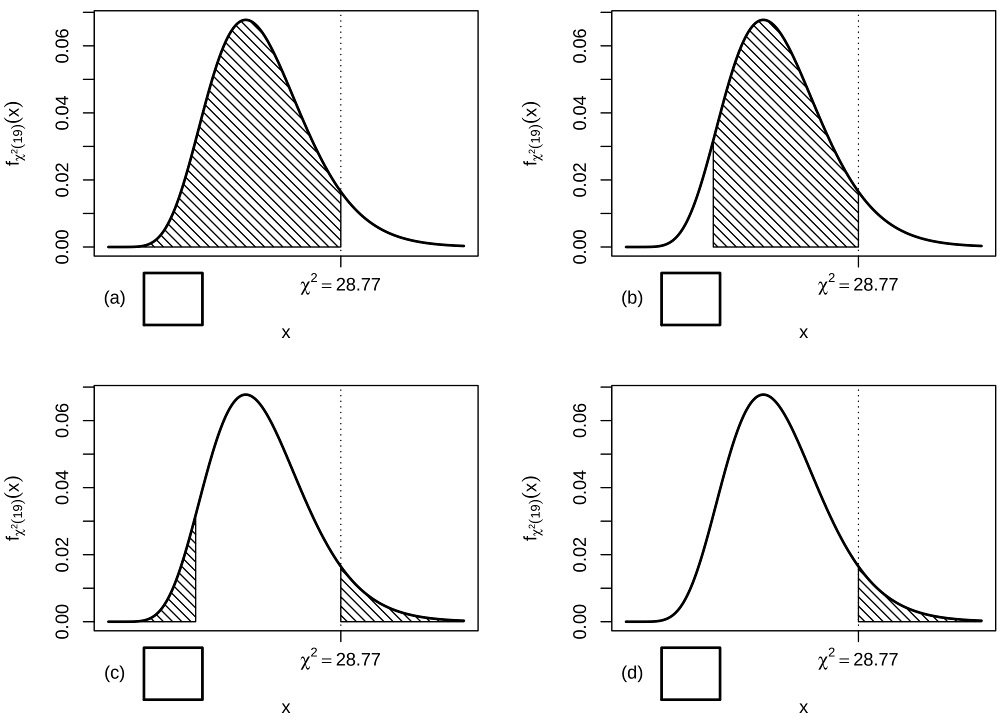
Aufgabe 2 (12 Punkte)

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. Sei $X_1, \dots, X_{20}$ 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 $n = 20$ soll

$$H_0 : \sigma^2 \leq \sigma_0^2 = 36 \qquad \text{gegen} \qquad H_1 : \sigma^2 > \sigma_0^2 = 36$$

mit einem Chi-Quadrat-Test getestet werden. Als realisierte Teststatistik erhält man $\chi^2 = 28.77$. 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.



2. Es sei $X_1, \dots, X_{64}$ eine einfache Stichprobe vom Umfang 64 zu Y mit $Y \sim N(59, 4^2)$. Dann gilt für die Teststatistik $N = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$ des Gauß-Tests für den Mittelwert einer normalverteilten Zufallsvariablen mit bekannter Varianz zur Nullhypothese $H_0 : \mu = 60$:

(a) $N \sim N(-2, 1)$ ☐

(b) $N \sim N(-2, 4^2)$ ☐

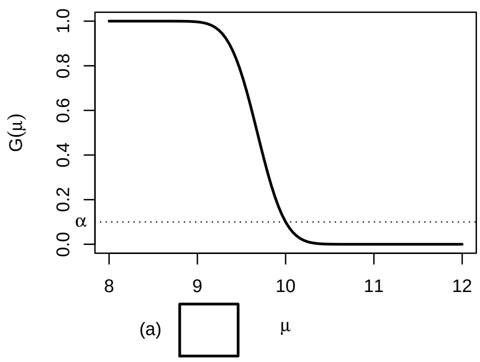
(c) $N \sim N(2, 1)$ ☐

(d) $N \sim N(2, 4^2)$ ☐

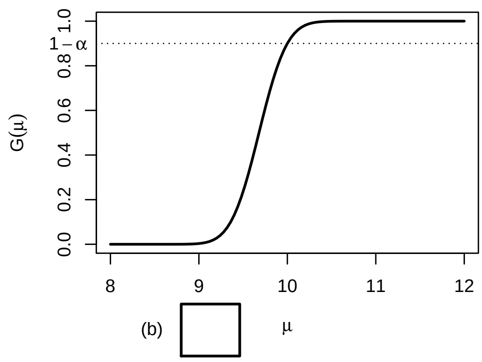
3. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{16}$ vom Umfang $n = 16$ zu einer $N(\mu, 1^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

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

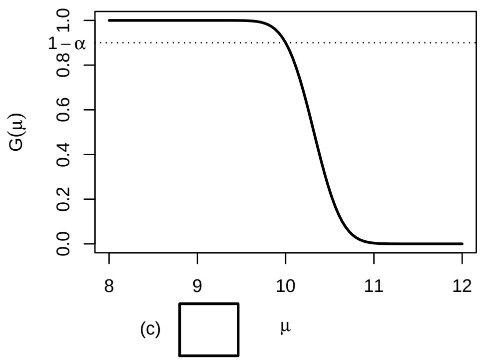
bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.
Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.



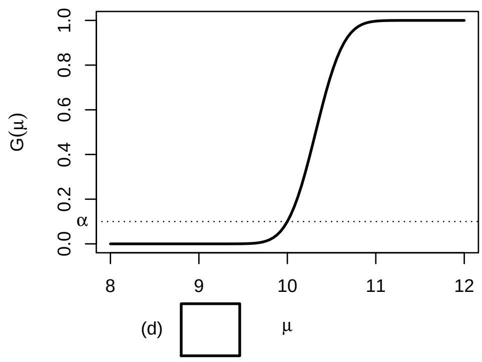
(a) ☐



(b) ☐



(c) ☐



(d) ☐

4. 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 α 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) $\chi^2 \in [0, \chi^2_{n-1;\frac{\alpha}{2}})$ ☐

(b) $\chi^2 \in (\chi^2_{n-1;1-\frac{\alpha}{2}}, \infty)$ ☐

(c) $\chi^2 \in [\chi^2_{n-1;\frac{\alpha}{2}}, \chi^2_{n-1;\alpha})$ ☐

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

Aufgabe 3 (4 Punkte)

In Abhängigkeit eines unbekannten Parameters $p > 0$ seien der Erwartungswert und die Varianz einer Zufallsvariablen Y gegeben durch $\mathrm{E}(Y) = \frac{3}{p}$ sowie $\mathrm{Var}(Y) = \frac{3}{p^2}$. Für $n \in \mathbb{N}$ sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y. Zeigen Sie, dass die Schätzfunktionen

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

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

Aufgabe 4 (3 + 1 + 6 = 10 Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $\lambda > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|\lambda) = \begin{cases} \frac{8}{9} \cdot \lambda^{-3} \cdot y^2 & \text{für } 0 \leq y \leq \frac{3}{2} \cdot \lambda \\ 0 & \text{sonst} \end{cases}$$

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

(a) Zeigen Sie, dass $\mathrm{E}(Y) = \frac{9}{8} \cdot \lambda$ gilt.

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

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

Hinweise:

- *Beachten Sie, dass Sie die Teile (b) und (c) auch ohne die Bearbeitung von Teil (a) 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.*

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

Eine Maschine produziert Bremsbeläge, deren Dicke erfahrungsgemäß normalverteilt mit einer Standardabweichung von $0.2[mm]$ um den tatsächlichen Erwartungswert schwankt. Die laufende Qualitätskontrolle soll eine Unterschreitung dieses Erwartungswerts gegenüber der mittleren Soll-Dicke $18[mm]$ mit Hilfe eines geeigneten statistischen Testverfahrens auf Basis der Realisation einer einfachen Stichprobe $x_1, \ldots, x_{16}$ aufdecken. Dabei darf eine derartige Unterschreitung nur mit einer Wahrscheinlichkeit von höchstens 5% fälschlicherweise signalisiert werden. Aus dem realisierten Stichprobenergebnis erhält man den Stichprobenmittelwert

$$\overline{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 17.871[mm] \ .$$

(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 16 keine Unterschreitung signalisieren, wenn der tatsächliche Erwartungswert der Dicke der Bremsbeläge $18.01[mm]$ beträgt?

(d) Wie groß muss der Stichprobenumfang bei der oben beschriebenen Qualitätskontrolle mindestens gewählt werden, wenn eine tatsächliche Unterschreitung des Erwartungswerts der Dicke um $0.15[mm]$ mit einer Wahrscheinlichkeit von mindestens 99% erkannt werden soll?

Aufgabe 6 (10 Punkte)

Zwei unterschiedlichen Gruppen mit 56 (Gruppe A) bzw. 46 (Gruppe B) Schmerzpatienten wird jeweils ein spezielles Schmerzmittel verabreicht. Nach einer festgelegten Zeit werden dann alle Schmerzpatienten gefragt, ob durch das verabreichte Medikament eine Linderung der Schmerzen eingetreten ist. In der Gruppe der Schmerzpatienten, denen Schmerzmittel A verabreicht wurde, beantworten 46 Personen diese Frage positiv, in der zu Schmerzmittel B gehörigen Gruppe 31 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 Schmerzmittel A besser wirkt als Schmerzmittel B (bezogen auf die Erfolgswahrscheinlichkeit für eine Linderung der Schmerzen). Formulieren Sie das Ergebnis auch in Form eines Antwortsatzes.

Aufgabe 7 (13 Punkte)

Um zu überprüfen, ob es einen Zusammenhang zwischen der Fachsemesteranzahl und dem Abschneiden in der Klausur (bestanden/nicht bestanden) gibt, hat der Dozent einer Statistik-Veranstaltung aus den Korrekturergebnissen der zugehörigen Klausuren aller 232 Teilnehmer die folgende Tabelle zusammengestellt:

	≤ 2 Fachsemester	≥ 3 Fachsemester
bestanden	155	24
nicht bestanden	47	6

Überprüfen Sie anhand dieses Datenmaterials zum Signifikanzniveau $\alpha = 0.05$, ob Fachsemesteranzahl und 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

Aufgabe 8 (14 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 Sommersemester 2018 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	134	68.21	694161	531.67
2	46	81.90	319667	247.04
3	52	88.76	424194	284.71

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.

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

$n \backslash m$	1	2	3	4	5	225	226	227	228	229
1	161.448	199.500	215.707	224.583	230.162	253.748	253.750	253.753	253.755	253.758
2	18.513	19.000	19.164	19.247	19.296	19.491	19.491	19.491	19.491	19.491
3	10.128	9.552	9.277	9.117	9.013	8.539	8.539	8.539	8.539	8.538
4	7.709	6.944	6.591	6.388	6.256	5.644	5.644	5.644	5.644	5.644
5	6.608	5.786	5.409	5.192	5.050	4.383	4.383	4.383	4.383	4.383
225	3.883	3.036	2.645	2.412	2.254	1.246	1.246	1.245	1.245	1.245
226	3.883	3.036	2.645	2.412	2.254	1.245	1.245	1.245	1.245	1.244
227	3.883	3.036	2.644	2.411	2.254	1.245	1.245	1.245	1.244	1.244
228	3.883	3.035	2.644	2.411	2.254	1.245	1.245	1.244	1.244	1.244
229	3.882	3.035	2.644	2.411	2.253	1.244	1.244	1.244	1.244	1.243

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

Zur Erklärung des Dieselpreises je Liter y_i (in Eurocent) durch den Erdölpreis je Barrel x_i (in US-Dollar) 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 insgesamt 11 Jahren wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

```
Residuals:
    Min      1Q  Median      3Q     Max 
-6.6542 -4.6908 -0.2338  4.4204  7.3969 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  35.4291     8.4385   4.199  0.00231 **
x             0.9796     0.4398   2.227  0.05292 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.393 on 9 degrees of freedom
Multiple R-squared:  0.3554,	Adjusted R-squared:  0.2837 
F-statistic: 4.961 on 1 and 9 DF,  p-value: 0.05292
```

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

(b) Welcher Anteil der Gesamtvarianz des Dieselpreises je Liter wird durch das lineare Modell erklärt?

(c) Entscheiden Sie mit Hilfe des zugehörigen p-Werts zum Signifikanzniveau $\alpha = 0.01$, ob β_1 signifikant von Null verschieden ist.

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

(e) Welchen Dieselpreis je Liter (in Eurocent) prognostiziert das Modell für ein Jahr mit einem Erdölpreis je Barrel von 90 (in US-Dollar)?

Aufgabe 10 (6 + 2 + 3 + 5 + 3 = 19 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 $n = 30$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{30} y_i = 17.182; \quad \sum_{i=1}^{30} y_i^2 = 44.889; \quad \sum_{i=1}^{30} x_i = 433.331;$$

$$\sum_{i=1}^{30} x_i^2 = 6555.723; \quad \sum_{i=1}^{30} x_i \cdot y_i = 197.697$$

(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.

10 Schließende Statistik SS 2019

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. Stimmen die Verteilungen von $X_1, \dots, X_n$ mit der Verteilung von Y überein, dann handelt es sich bei $X_1, \dots, X_n$ stets um eine einfache Stichprobe vom Umfang n zu Y.	☐	☐
2. Für Schätzwerte $\widehat{\theta}_{\text{ML}}$ nach der Maximum-Likelihood-Methode und Schätzwerte $\widehat{\theta}_{\text{MM}}$ nach der Momenten-Methode (basierend auf derselben Stichprobenrealisation) gilt stets $\widehat{\theta}_{\text{ML}} \geq \widehat{\theta}_{\text{MM}}$.	☐	☐
3. Bei der Anwendung der Maximum-Likelihood-Methode für stetige Verteilungsfamilien kann die Likelihoodfunktion wegen der (vorausgesetzten) Unabhängigkeit der Stichprobenzufallsvariablen als Produkt von Randdichten dargestellt werden.	☐	☐
4. Sind $\widehat{\theta}$ und $\widetilde{\theta}$ zwei Schätzfunktionen zur Schätzung eines Parameters $\theta \in \Theta$ und ist $\widehat{\theta}$ wirksamer als $\widetilde{\theta}$, so gilt $\text{Var}(\widehat{\theta}) < \text{Var}(\widetilde{\theta})$ für alle $\theta \in \Theta$.	☐	☐
5. Bei einem zweiseitigen Gauß-Test für den Mittelwert bei bekannter Varianz ist der p-Wert umso größer, je näher $\overline{X}$ am hypothetischen Erwartungswert μ_0 liegt.	☐	☐
6. Die Gütefunktion eines zweiseitigen Gauß-Tests gibt zu tatsächlichen Erwartungswerten $\mu \neq \mu_0$ an, mit welcher Wahrscheinlichkeit die Durchführung des Tests zu einer falschen Entscheidung führt.	☐	☐
7. Lehnt ein Chi-Quadrat-Anpassungstest die Nullhypothese zu einem Signifikanzniveau von $\alpha = 0.10$ ab, so wird die Nullhypothese stets auch bei einem entsprechenden Test zum Signifikanzniveau von $\widetilde{\alpha} = 0.05$ verworfen.	☐	☐
8. Im einfachen linearen Regressionsmodell $y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2),$ liegt der Punkt $(\overline{x}, \overline{y})$ stets auf der nach der KQ-Methode bestimmten Regressionsgeraden.	☐	☐

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 bearbei-

tet!).

1. Es sei $X_1, \dots, X_{25}$ eine einfache Stichprobe vom Umfang 25 zu Y mit $Y \sim N(100, 20^2)$. Dann gilt für die Verteilung von $\overline{X} = \frac{1}{25}\sum_{i=1}^{25} X_i$:

   (a) $\overline{X} \sim N\left(100, 2^2\right)$

   (b) $\overline{X} \sim N\left(100, 4^2\right)$

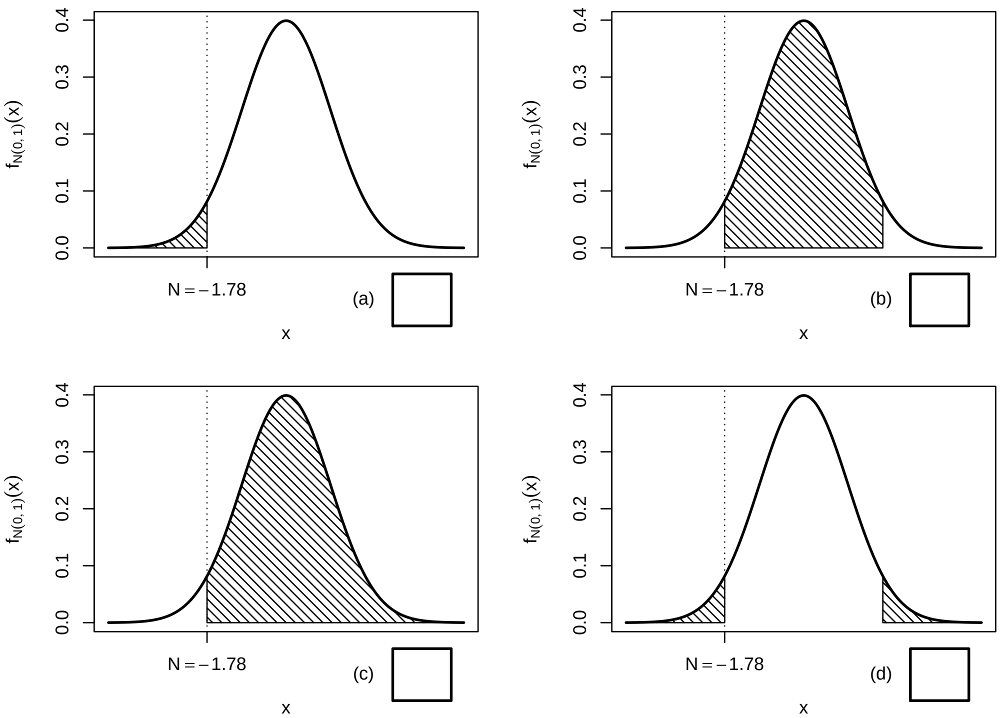
   (c) $\overline{X} \sim N\left(100, 10^2\right)$

   (d) $\overline{X} \sim N\left(100, 20^2\right)$

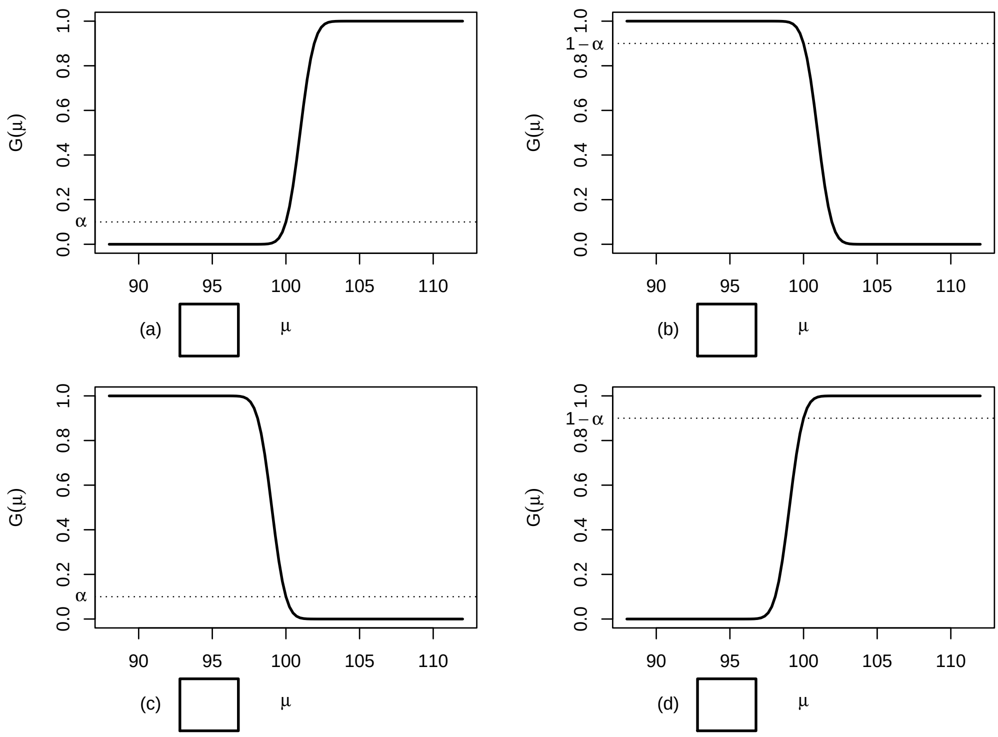
2. Sei $X_1, \dots, X_{36}$ eine einfache Stichprobe zu einer $N(\mu, \sigma_0^2)$-verteilten Zufallsvariablen Y mit unbekanntem Parameter μ und bekanntem $\sigma_0^2 = 4^2$. Auf der Grundlage einer Stichprobenrealisation zu dieser einfachen Stichprobe vom Umfang $n = 36$ soll

   $$H_0 : \mu = \mu_0 = 40 \qquad \text{gegen} \qquad H_1 : \mu \neq \mu_0 = 40$$

   mit einem Gauß-Test getestet werden. Als realisierte Teststatistik erhält man $N = -1.78$. 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 $X_1, \dots, X_{64}$ vom Umfang $n = 64$ zu einer $N(\mu, 6^2)$-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.



4. Bei der Durchführung eines linksseitigen t-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei unbekannter Varianz (mit $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$) wird die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ abgelehnt. Dann gilt für die Testentscheidungen des rechtsseitigen (mit $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$) und des zweiseitigen (mit $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$) Tests zum **veränderten Signifikanzniveau** $\alpha = 0.10$:
   (a) Der rechtsseitige Test lehnt H_0 nicht ab, der zweiseitige Test lehnt H_0 ab. ☐
   (b) Der rechtsseitige Test lehnt H_0 nicht ab, die Entscheidung des zweiseitigen Tests kann für oder gegen H_0 ausfallen. ☐
   (c) Der rechtsseitige Test lehnt H_0 ab, die Entscheidung des zweiseitigen Tests kann für oder gegen H_0 ausfallen. ☐
   (d) Sowohl beim rechtsseitigen als auch beim zweiseitigen Test kann die Entscheidung für oder gegen H_0 ausfallen. ☐

Aufgabe 3 (4 Punkte)

Für $\lambda > 0$ sei $Y \sim \text{Exp}(\lambda)$, es gilt also insbesondere $\text{E}(Y) = \frac{1}{\lambda}$ sowie $\text{Var}(Y) = \frac{1}{\lambda^2}$. $X_1, \dots, X_n$ sei für $n \in \mathbb{N}$ eine einfache Stichprobe vom Umfang n zu Y. Untersuchen Sie, ob die Schätzfunktionen
$$T_n(X_1, \dots, X_n) := \frac{1}{2n} \sum_{i=1}^{n} X_i^2$$

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

Aufgabe 4 $(3 + 1 + 6 = 10$ Punkte)

Die Verteilung einer Zufallsvariablen Y sei in Abhängigkeit des unbekannten Parameters $a > 0$ durch die folgende Dichtefunktion gegeben:

$$f_Y(y|a) = \begin{cases} \frac{1}{4} \cdot a^{-4} \cdot y^3 & \text{für } 0 \leq y \leq 2 \cdot a \\ 0 & \text{sonst} \end{cases}$$

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

(a) Zeigen Sie, dass $\mathrm{E}(Y) = \frac{8}{5} \cdot a$ gilt.

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

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

Hinweise:

- *Beachten Sie, dass Sie die Teile (b) und (c) auch ohne die Bearbeitung von Teil (a) 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.*

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

Bei der Abfüllung von Babynahrung weiß der Hersteller aus Erfahrung, dass die verwendete Maschine eine Standardabweichung von $2[g]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Maschine im Mittel weniger als die auf dem Produkt ausgezeichneten $190[g]$ in die Gläser einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 16 Gläser entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{16}$ als Realisation einer einfachen Stichprobe vom Umfang 16 zur annahmegemäß $N(\mu, 2^2[g^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 188.959[g] .$$

(a) 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.

(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) Wie groß ist die Wahrscheinlichkeit, bei der Ziehung einer Stichprobe und der anschließenden Durchführung des Tests aus Teil (a) eine Testentscheidung zu Gunsten der Nullhypothese zu erhalten, falls die tatsächliche mittlere Abfüllmenge $188[g]$ beträgt?

Aufgabe 6 (9 + 7 = 16 Punkte)

Ein Hersteller von Metallwaren produziert unter anderem Stahlstifte mit einer Soll-Länge von 8 [cm]. Es soll angenommen werden, dass die Länge der hergestellten Stahlstifte gemäß einer Normalverteilung mit unbekanntem Erwartungswert und unbekannter Varianz schwankt. Zur Überprüfung der Qualität der produzierten Stahlstifte werden 10 Stahlstifte aus dem laufenden Produktionsprozess entnommen. Die gemessenen Längen

$$7.94, 8.10, 8.03, 8.01, 8.16, 8.04, 8.14, 7.99, 8.03, 8.04$$

seien als Realisation einer einfachen Stichprobe zur normalverteilten Grundgesamtheit aufzufassen. Aus dieser Realisation wurde bereits $s^2 = 0.004551$ [cm^2] berechnet.

(a) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die tatsächliche mittlere Länge der produzierten Stahlstifte im Vergleich zur angegebenen Soll-Länge von 8 [cm] zu groß ist. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die Varianz der Länge der produzierten Stahlstifte von der vom Hersteller angegebenen Toleranz $\sigma_0^2 = 0.01$ [cm^2] abweicht. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) 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
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

Aufgabe 7 (7 Punkte)

Zur Beurteilung der Präzision zweier Messgeräte A und B wird eine Referenzgröße jeweils unabhängig voneinander mit beiden Messgeräten wiederum jeweils unabhängig voneinander mehrfach gemessen. Es werde angenommen, dass die gemessenen Werte Y^A bzw. Y^B der beiden Messgeräte jeweils normalverteilt seien mit unbekannten Erwartungswerten μ_A bzw. μ_B sowie unbekannten Varianzen σ_A^2 bzw. σ_B^2. Die Ergebnisse der wiederholten Messungen lassen sich als (voneinander unabhängige) einfache Stichproben $X_1^A, \ldots, X_{17}^A$ vom Umfang 17 zu Y^A sowie $X_1^B, \ldots, X_{19}^B$ vom Umfang 19 zu Y^B auffassen, aus den zugehörigen Realisationen wurden bereits die Mittelwerte $\overline{x^A} = 100.167$ bzw. $\overline{x^B} = 99.456$ sowie die Stichprobenvarianzen $s_{Y^A}^2 = 0.762$ bzw. $s_{Y^B}^2 = 1.847$ berechnet. Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob Messgerät A eine höhere Präzision (im Sinne einer geringeren Streuung) als Messgerät B hat. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m,n)$*-Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n\backslash m$	11	12	13	14	15	16	17	18	19	20
11	2.818	2.788	2.761	2.739	2.719	2.701	2.685	2.671	2.658	2.646
12	2.717	2.687	2.660	2.637	2.617	2.599	2.583	2.568	2.555	2.544
13	2.635	2.604	2.577	2.554	2.533	2.515	2.499	2.484	2.471	2.459
14	2.565	2.534	2.507	2.484	2.463	2.445	2.428	2.413	2.400	2.388
15	2.507	2.475	2.448	2.424	2.403	2.385	2.368	2.353	2.340	2.328
16	2.456	2.425	2.397	2.373	2.352	2.333	2.317	2.302	2.288	2.276
17	2.413	2.381	2.353	2.329	2.308	2.289	2.272	2.257	2.243	2.230
18	2.374	2.342	2.314	2.290	2.269	2.250	2.233	2.217	2.203	2.191
19	2.340	2.308	2.280	2.256	2.234	2.215	2.198	2.182	2.168	2.155
20	2.310	2.278	2.250	2.225	2.203	2.184	2.167	2.151	2.137	2.124

Aufgabe 8 (14 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 2017/18 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	87	81.09	615649	506.66
2	24	95.38	221282	128.08
3	51	99.29	509830	140.93

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.

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

$n\backslash m$	1	2	3	4	5	155	156	157	158	159
1	161.448	199.500	215.707	224.583	230.162	253.492	253.497	253.503	253.508	253.513
2	18.513	19.000	19.164	19.247	19.296	19.489	19.489	19.489	19.489	19.489
3	10.128	9.552	9.277	9.117	9.013	8.544	8.544	8.544	8.544	8.544
4	7.709	6.944	6.591	6.388	6.256	5.651	5.651	5.651	5.651	5.651
5	6.608	5.786	5.409	5.192	5.050	4.391	4.391	4.391	4.390	4.390
155	3.902	3.054	2.663	2.430	2.273	1.303	1.303	1.303	1.302	1.302
156	3.902	3.054	2.663	2.430	2.272	1.303	1.302	1.302	1.301	1.301
157	3.901	3.054	2.662	2.429	2.272	1.302	1.302	1.301	1.301	1.300
158	3.901	3.053	2.662	2.429	2.271	1.302	1.301	1.301	1.300	1.300
159	3.901	3.053	2.661	2.429	2.271	1.301	1.300	1.300	1.300	1.299

Aufgabe 9 (1 + 1 + 1 + 1 + 1 + 1 + 1 = 7 Punkte)

Zur Erklärung der stetigen Wochenrenditen der BASF-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 
-2.0828 -0.6635 -0.2124  0.7805  2.1818 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   0.5098     0.3602   1.415   0.1789  
x             0.7064     0.2473   2.856   0.0127 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.313 on 14 degrees of freedom
Multiple R-squared:  0.3682,	Adjusted R-squared:  0.323 
F-statistic: 8.158 on 1 and 14 DF,  p-value: 0.01269
```

(a) Wie viele Wochenrenditen 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 stetigen Wochenrenditen der BASF-Aktie wird durch das lineare Modell erklärt?

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

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

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

Aufgabe 10 (6 + 2 + 2 + 3 + 3 + 5 = 21 Punkte)

Zur Schätzung eines einfachen linearen Regressionsmodells

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

aus einer Stichprobe vom Umfang $n = 16$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{16} y_i = 464.89; \quad \sum_{i=1}^{16} y_i^2 = 13947.79; \quad \sum_{i=1}^{16} x_i = 91.27;$$

$$\sum_{i=1}^{16} x_i^2 = 570.42; \quad \sum_{i=1}^{16} x_i \cdot y_i = 2792.65$$

(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) Geben Sie ein symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_2 an.

(f) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für y_0 gegeben $x_0 = 6$ an.

11 Schließende Statistik WS 2019/20

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. Die Zufallsvariablen $X_1, \ldots, X_n$ einer einfachen Stichprobe vom Umfang n zu Y seien normalverteilt. Dann ist auch Y stets normalverteilt.	☐	☐
2. Eine Familie von im quadratischen Mittel konsistenten Schätzfunktionen ist zumindest asymptotisch erwartungstreu.	☐	☐
3. Lehnt ein linksseitiger Chi-Quadrat-Test für die Varianz einer normalverteilten Zufallsvariablen bei unbekanntem Erwartungswert die Nullhypothese zu einem Signifikanzniveau von $\alpha = 0.05$ ab, so wird die Nullhypothese stets auch bei einem entsprechenden Test zum Signifikanzniveau von $\widetilde{\alpha} = 0.10$ verworfen.	☐	☐
4. Im Rahmen einer statistischen Qualitätskontrolle mit Hilfe eines zweiseitigen Gauß-Tests zum Signifikanzniveau $\alpha = 0.05$ wird man – auch wenn alle Annahmen des Tests erfüllt sind – durchschnittlich in einer von 20 Kontrollen die Nullhypothese ablehnen, obwohl sie tatsächlich erfüllt ist.	☐	☐
5. Beim zweiseitigen Gauß-Test für den Erwartungswert bei bekannter Varianz (mit $H_0 : \mu = \mu_0$) ist der p-Wert stets umso kleiner, je weiter der Erwartungswert μ von μ_0 entfernt ist.	☐	☐
6. Zur Durchführung eines Mittelwertvergleichs auf Basis einer zweidimensional normalverteilten (verbundenen) Stichprobe mit einem t-Differenzentest liegen als Stichprobeninformation 11 Paare (mit jeweils zwei Beobachtungswerten) vor. Dann ist zur Konstruktion des kritischen Bereichs eine t-Verteilung mit 20 Freiheitsgraden zu verwenden.	☐	☐
7. Die Teststatistik F der einfachen Varianzanalyse kann als Quotient von zwei Größen verstanden werden, die bei Gültigkeit von H_0 beide sinnvolle Schätzer für die unbekannte Varianz σ^2 sind.	☐	☐
8. Im einfachen linearen Regressionsmodell $$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, n,$$ gilt für die Summe der quadrierten Residuen stets $\sum_{i=1}^n \widehat{u}_i^2 = 0$ (mit $\widehat{u}_i = y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_i)$).	☐	☐

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. Konfidenzintervalle für den Erwartungswert einer normalverteilten Zufallsvariablen mit bekannter Varianz sind umso breiter,

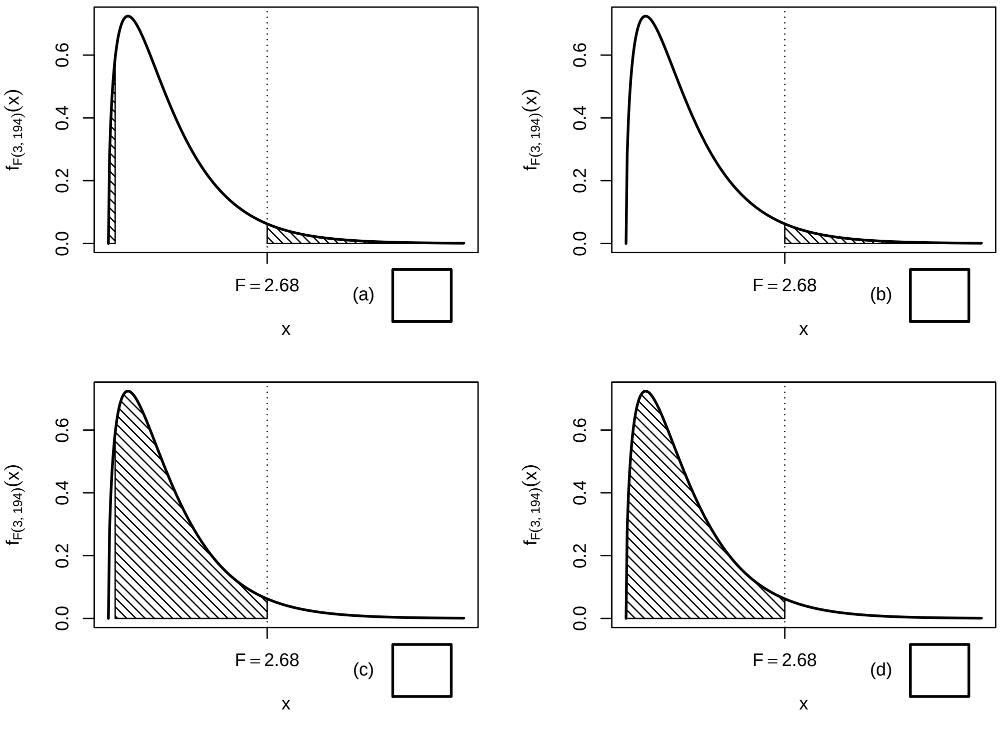
   (a) je kleiner die Varianz und je kleiner das Konfidenzniveau $1-\alpha$ ist. ☐

   (b) je kleiner die Varianz und je größer das Konfidenzniveau $1-\alpha$ ist. ☐

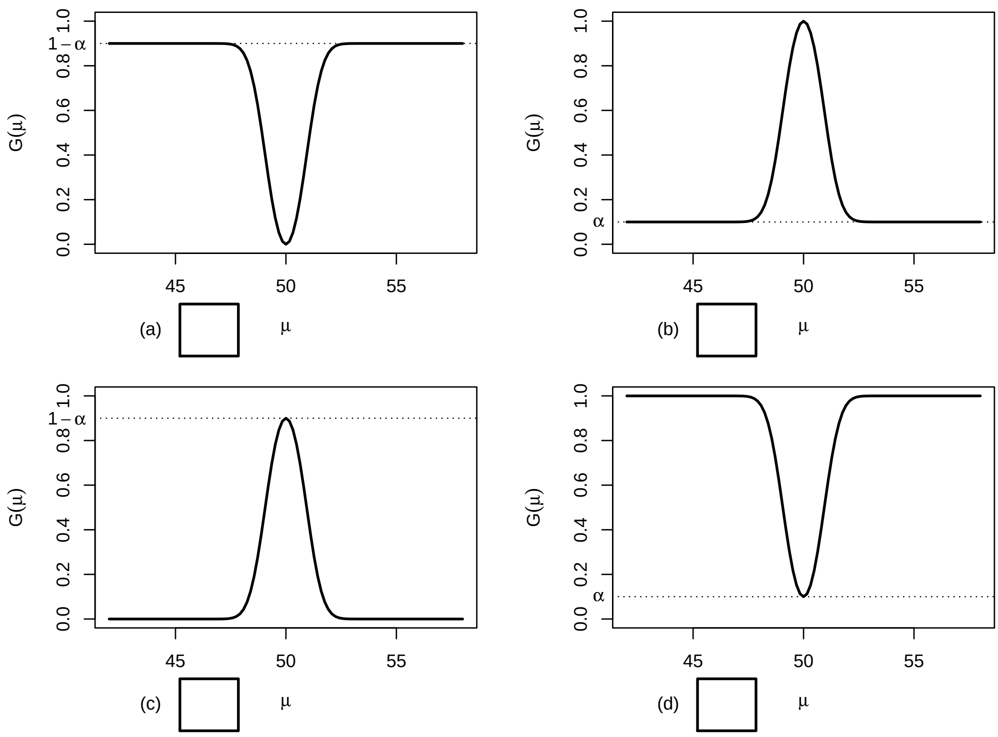
   (c) je größer die Varianz und je kleiner das Konfidenzniveau $1-\alpha$ ist. ☐

   (d) je größer die Varianz und je größer das Konfidenzniveau $1-\alpha$ ist. ☐

2. Bei der Durchführung einer einfachen Varianzanalyse mit $k = 4$ Faktorstufen und einem Gesamtstichprobenumfang von $n = 198$ erhält man die realisierte Teststatistik $F = 2.68$. 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. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{49}$ vom Umfang $n = 49$ zu einer $N(\mu, 4^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen
$$H_0 : \mu = 50 \qquad \text{gegen} \qquad H_1 : \mu \neq 50$$
bei einem Signifikanzniveau von $\alpha = 0.1$ betrachtet.
Markieren Sie die Abbildung, welche die Gütefunktion des oben genannten Tests korrekt darstellt.



4. Als p-Wert zur realisierten Teststatistik eines zweiseitigen t-Tests für den Mittelwert einer normalverteilten Grundgesamtheit bei unbekannter Varianz (mit $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$) erhält man $p = 0.0524$. Außerdem ist bekannt, dass die Teststatistik positiv ist. Dann gilt für die p-Werte des linksseitigen Tests (mit $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$) bzw. des rechtsseitigen Tests (mit $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$):
   (a) Der p-Wert des linksseitigen Tests beträgt 0.1048, der p-Wert des rechtsseitigen Tests 0.8952. ☐
   (b) Der p-Wert des linksseitigen Tests beträgt 0.8952, der p-Wert des rechtsseitigen Tests 0.1048. ☐
   (c) Der p-Wert des linksseitigen Tests beträgt 0.0262, der p-Wert des rechtsseitigen Tests 0.9738. ☐
   (d) Der p-Wert des linksseitigen Tests beträgt 0.9738, der p-Wert des rechtsseitigen Tests 0.0262. ☐

Aufgabe 3 $(3 + 3 + 1 = 7$ Punkte)

In Abhängigkeit eines unbekannten Parameters p mit $0 < p < 1$ sei die Verteilung einer Zufallsvariablen Y gegeben durch:
Für $n \in \mathbb{N}$ sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y.

(a) Zeigen Sie, dass $\mathrm{E}(Y) = \dfrac{5 - p}{2}$ gilt.

y_i	1	2	4
$p_Y(y_i)$	$\frac{1-p}{2}$	p	$\frac{1-p}{2}$

(b) Zeigen Sie, dass die Schätzfunktionen

$$T_n(X_1, \ldots, X_n) := 5 - 2\overline{X}$$

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

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

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

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

$$f_Y(y|\lambda) = \begin{cases} \dfrac{\lambda \cdot 3^\lambda}{y^{\lambda+1}} & \text{für } y \geq 3 \\ 0 & \text{sonst} \end{cases}$$

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

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

(b) Zeigen Sie, dass $E(Y) = \dfrac{3\lambda}{\lambda - 1}$ gilt.

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

Aufgabe 5 (7 + 3 + 3 = 13 Punkte)

Bei der Herstellung von gerösteten Kaffeebohnen weiß der Hersteller aus Erfahrung, dass die verwendete Abfüllanlage eine Standardabweichung von $10[g]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Abfüllanlage im Mittel mehr als die auf dem Produkt ausgezeichneten $1000[g]$ in die Packungen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 25 Packungen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{25}$ als Realisation einer einfachen Stichprobe vom Umfang 25 zur annahmegemäß $N(\mu, 10^2[g^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 1002.913[g] \,.$$

(a) 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.

(b) Wie groß ist die Wahrscheinlichkeit, bei der Ziehung einer Stichprobe und der anschließenden Durchführung des Tests aus Teil (a) eine Testentscheidung zu Gunsten der Nullhypothese zu erhalten, falls die tatsächliche mittlere Abfüllmenge $1005[g]$ beträgt?

(c) Erläutern Sie kurz – wahlweise mit Hilfe der Gütefunktion oder der Verteilung der Teststatistik in Abhängigkeit von μ –, warum die Fehlerwahrscheinlichkeit 1. Art mit fallendem μ fällt.

Aufgabe 6 (10 + 8 = 18 Punkte)

Um zu überprüfen, ob zwei Powerbank-Fabrikate („Typ A" bzw. „Typ B") mit nominell übereinstimmender Kapazität tatsächlich eine im Durchschnitt übereinstimmende entnehmbare Energie aufweisen, soll ein statistischer Test durchgeführt werden. Hierbei soll davon ausgegangen werden, dass die jeweils entnehmbare Energie Y^A bzw. Y^B normalverteilt sei mit unbekanntem Erwartungswert μ_A bzw. μ_B und unbekannter Varianz σ_A^2 bzw. σ_B^2. Es soll überprüft werden, ob Powerbanks vom Typ B im Mittel eine höhere entnehmbare Energie als solche vom Typ A besitzen.
Aus einer Kapazitätsmessung mit $n_A = 9$ Exemplaren der Powerbank vom Typ A sowie $n_B = 11$ Exemplaren der Powerbank vom Typ B erhielt man Realisationen jeweils voneinander unabhängiger einfacher Stichproben $X_1^A, \ldots, X_9^A$ zu Y^A sowie $X_1^B, \ldots, X_{11}^B$ zu Y^B und hieraus die zugehörigen Mittelwerte $\overline{x^A} = 9854$ bzw. $\overline{x^B} = 9964$ sowie die Stichproben**standardabweichungen** $s_{Y^A} = 156$ bzw. $s_{Y^B} = 170$.

(a) Testen Sie unter der Annahme $\sigma_A^2 = \sigma_B^2$ zum Signifikanzniveau $\alpha = 0.05$ die Hypothese, dass Powerbanks vom Typ B im Mittel eine höhere entnehmbare Energie als solche vom Typ A besitzen. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Überprüfen Sie mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.10$, ob die in Teil (a) getroffene Annahme der Varianzgleichheit auf Grundlage der vorhandenen Stichprobeninformation verworfen werden muss. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (b) den folgenden Tabellenausschnitt mit 0.95*-Quantilen von* $F(m, n)$*-Verteilungen sowie ggf. die Rechenregel* $F_{m,n;p} = \frac{1}{F_{n,m;1-p}}$.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

Aufgabe 7 (11 + 3 = 14 Punkte)

Mit einem Chi-Quadrat-Anpassungstest soll zum Signifikanzniveau $\alpha = 0.05$ überprüft werden, ob man bei einem beobachteten Stichprobenergebnis von der Realisation einer einfachen Stichprobe vom Umfang $n = 200$ zu einer Geom(0.5)-verteilten Zufallsvariablen ausgehen kann. Die Stichprobeninformation liege in Form der folgenden Häufigkeitsverteilung vor:

a_i	0	1	2	≥ 3
n_i	71	58	33	38

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

(b) Zum Test, ob die angegebene Häufigkeitsverteilung als Stichprobenrealisation zu *irgendeiner* geometrischen Verteilung Geom(p) (für ein beliebiges $p \in (0,1)$) plausibel ist, wurde der Verteilungsparameter p mit Hilfe einer ML-Schätzung aus den wie oben klassierten Daten (zu $\widehat{p} = 0.405$) geschätzt und damit die (neue) Teststatistik $\chi^2 = 4.2808$ berechnet. Zu welchem Ergebnis kommt dieser Test? Begründen Sie Ihre Antwort durch die Angabe des zugehörigen kritischen Bereichs.

Hinweise:

- *Die geometrische Verteilung mit Parameter* $p = 0.5$ *hat den Träger* $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$ *und die Wahrscheinlichkeitsfunktion*

$$p_{\text{Geom}(0.5)} : \mathbb{N}_0 \to [0,1]; p_{\text{Geom}(0.5)}(i) = (1-0.5)^i \cdot 0.5 \; .$$

- *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

Aufgabe 8 (1 + 1 + 1 + 2 + 1 = 6 Punkte)

Zur Erklärung des Kraftstoffverbrauchs y_i (in [l]) durch die zurückgelegte Distanz x_i (in [km]) 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 aus mehreren Tankvorgängen wurde das lineare Modell mit der Statistik-Software R wie folgt geschätzt:

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

Residuals:
    Min      1Q  Median      3Q     Max
-3.6999 -1.3933 -0.1942  1.6728  4.1259

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.402311   1.487031   1.616    0.109
x           0.051385   0.001966  26.135   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.01 on 120 degrees of freedom
Multiple R-squared:  0.8506,       Adjusted R-squared:  0.8493
F-statistic: 683.1 on 1 and 120 DF,  p-value: < 2.2e-16
```

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

(b) Welcher Anteil der Gesamtvarianz des Kraftstoffverbrauchs wird durch das lineare Modell erklärt?

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

(d) Geben Sie ein Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_2 an.

(e) Welchen Kraftstoffverbrauch (in [l]) prognostiziert das Modell für eine zurückgelegte Distanz von 723 (in [km])?

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 $n = 25$ wurden bereits die folgenden Zwischenwerte errechnet:

$$\sum_{i=1}^{25} y_i = 65.314; \quad \sum_{i=1}^{25} y_i^2 = 191.836; \quad \sum_{i=1}^{25} x_i = 82.292;$$

$$\sum_{i=1}^{25} x_i^2 = 298.149; \quad \sum_{i=1}^{25} x_i \cdot y_i = 202.447$$

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

(f) Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für $\mathrm{E}(y_0)$ gegeben $x_0 = 4$ an.

12 Schließende Statistik SS 2020

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) $F(2, 90)$ ☐

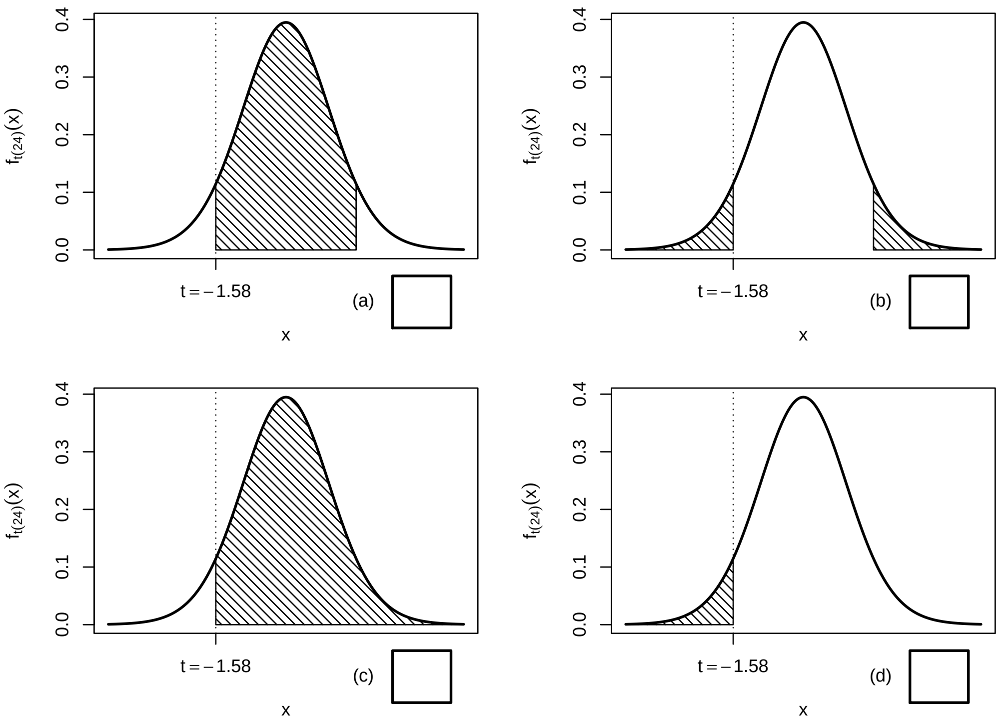
   (c) $F(3, 87)$ ☐

   (d) $F(2, 87)$ ☐

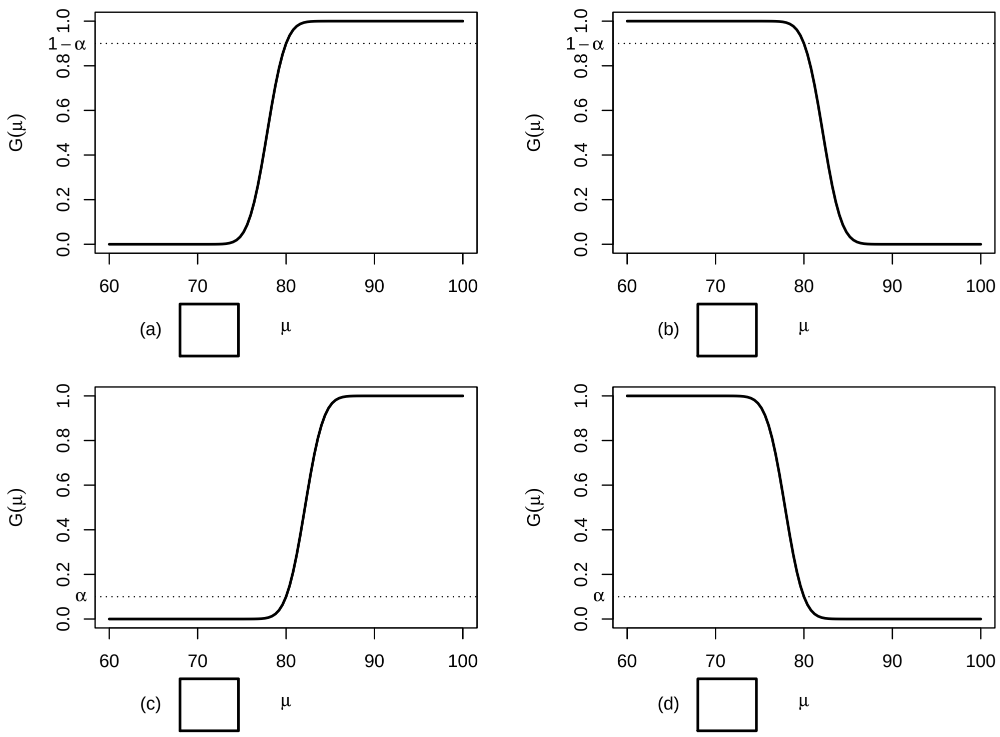
2. Sei $X_1, \ldots, X_{25}$ eine einfache Stichprobe zu einer $N(\mu, \sigma^2)$-verteilten Zufallsvariablen Y mit unbekannten Parametern μ und σ^2. Auf der Grundlage einer Stichprobenrealisation zu dieser einfachen Stichprobe vom Umfang $n = 25$ soll

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

   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!)*

Aufgabe 4 (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.*

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 Maschine 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?

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.

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.

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

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.

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.

13 Schließende Statistik WS 2020/21

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 stochastisch unabhängig.	☐	☐
2. Ist eine Familie von Schätzfunktionen $T_n, n \in \mathbb{N}$, konsistent im quadratischen Mittel für einen Parameter $\theta \in \mathbb{R}$, so ist zumindest eine dieser Schätzfunktionen auch effizient in der Klasse der für θ erwartungstreuen Schätzfunktionen.	☐	☐
3. Die Breite von Konfidenzintervallen für den Erwartungswert μ einer normalverteilten Zufallsvariablen bei bekannter Varianz verkleinert sich mit wachsendem Stichprobenumfang.	☐	☐
4. Lehnt man in einem statistischen Test die Nullhypothese H_0 auf Grundlage eines p-Werts der realisierten Teststatistik von $p = 0.0286$ ab, so gilt H_0 nur mit einer Wahrscheinlichkeit von maximal 2.86%.	☐	☐
5. 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 gilt für die Realisation t der Teststatistik: $t \in [-t_{n-1,1-\frac{\alpha}{2}}, -t_{n-1,1-\alpha})$	☐	☐
6. Mit einem Chi-Quadrat-Anpassungstest soll anhand einer einfachen Stichprobe vom Umfang $n = 100$ überprüft werden, ob die Grundgesamtheit geometrisch verteilt ist. Nach Wahl einer geeigneten Klassierung aus 6 Klassen wird dazu zunächst der unbekannte Parameter der geometrischen Verteilung durch eine ML-Schätzung aus den klassierten Daten ermittelt. Damit ist zur Konstruktion des Ablehnungsbereiches die χ^2-Verteilung mit 4 Freiheitsgraden zu verwenden.	☐	☐
7. Beim Test zum Vergleich von zwei Anteilswerten (als Spezialfall des 2-Stichproben-t-Tests) ergibt sich unter der speziellen Annahme $p_A = p_B$ die Voraussetzung $\sigma_A^2 = \sigma_B^2$ von selbst.	☐	☐
8. Die Anwendung der Kleinst-Quadrate-Methode bedeutet, die Summe der quadrierten horizontalen Abstände der Beobachtungspunkte zur Regressionsgeraden zu minimieren.	☐	☐

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!).

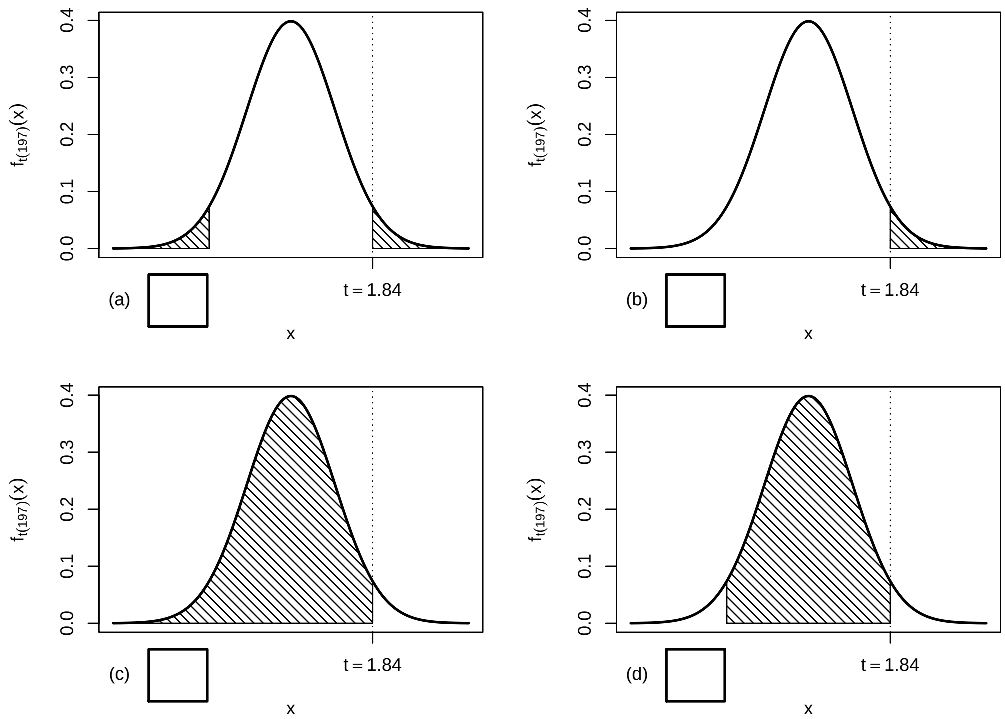
1. Es sei $X_1, \ldots, X_{25}$ eine einfache Stichprobe vom Umfang 25 zu Y mit $Y \sim N(33, 5^2)$. Dann gilt für die Teststatistik $N = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$ des Gauß-Tests für den Mittelwert einer normalverteilten Zufallsvariablen mit bekannter Varianz zur Nullhypothese $H_0 : \mu = 30$:

   (a) $N \sim N\left(-3, 5^2\right)$ ☐
   (b) $N \sim N\left(3, 5^2\right)$ ☐
   (c) $N \sim N\left(-3, 1\right)$ ☐
   (d) $N \sim N\left(3, 1\right)$ ☐

2. Sei $X_1, \ldots, X_{198}$ 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 $n = 198$ soll

$$H_0 : \mu \leq \mu_0 = 250 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0 = 250$$

   mit einem t-Test getestet werden. Als realisierte Teststatistik erhält man $t = 1.84$. 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. Wird die Teststatistik der einfachen Varianzanalyse als Quotient mit dem Zähler $SB/(k-1)$ und dem Nenner $SW/(n-k)$ notiert und bezeichnet σ^2 die Varianz der Zufallsvariablen $Y_1, \ldots, Y_k$, so

   (a) sind Zähler und Nenner stets sinnvolle Schätzer für σ^2. ☐

   (b) ist der Zähler stets, der Nenner nur unter H_0 ein sinnvoller Schätzer für σ^2. ☐

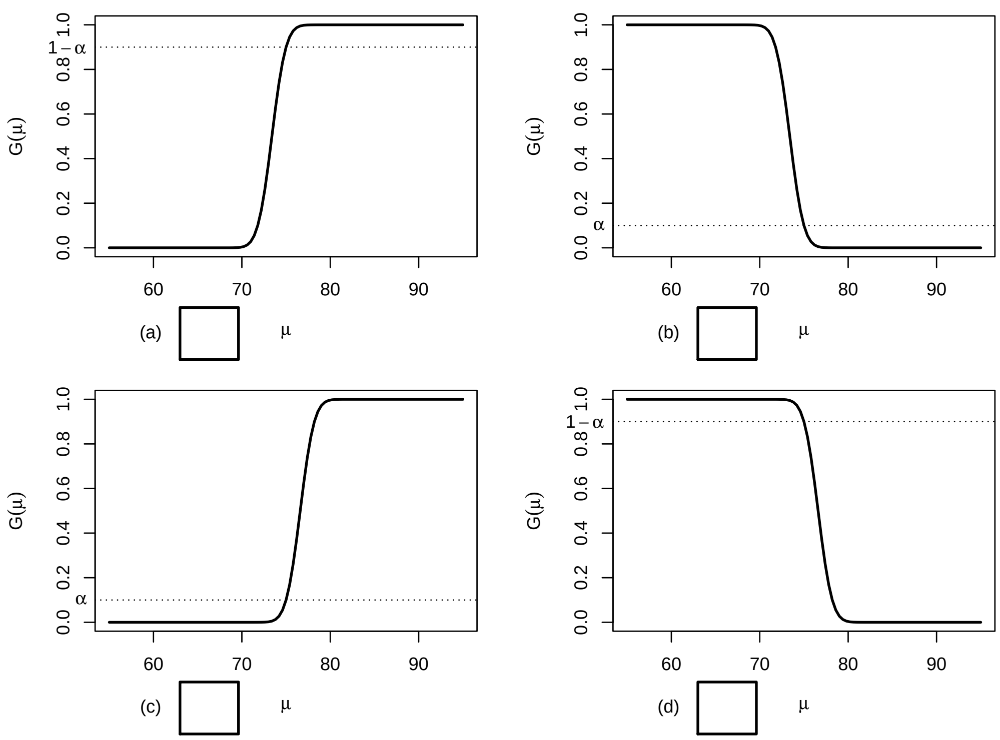
   (c) ist der Zähler nur unter H_0, der Nenner stets ein sinnvoller Schätzer für σ^2. ☐

   (d) sind Zähler und Nenner nur unter H_0 sinnvolle Schätzer für σ^2. ☐

4. Auf der Grundlage einer einfachen Stichprobe $X_1, \ldots, X_{64}$ vom Umfang $n = 64$ zu einer $N(\mu, 10^2)$-verteilten Zufallsvariablen wird ein Gauß-Test zur Überprüfung der Hypothesen

$$H_0 : \mu \leq 75 \qquad \text{gegen} \qquad H_1 : \mu > 75$$

   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 + 2 = 8 Punkte)

Für $\sigma > 0$ sei die Zufallsvariable Y Rayleigh-verteilt mit Parameter σ. Es gilt dann $\mathrm{E}(Y) = \sqrt{\frac{\pi}{2}}\sigma$ sowie $\mathrm{Var}(Y) = \frac{4-\pi}{2}\sigma^2$. $X_1, \ldots, X_n$ sei für $n \in \mathbb{N}$ eine einfache Stichprobe vom Umfang n zu Y.

(a) Zeigen Sie, dass die Schätzfunktionen

$$T_n(X_1, \ldots, X_n) := \frac{1}{3n} \sum_{i=1}^{n} X_i^2$$

nicht erwartungstreu für den **quadrierten** Parameter σ^2 sind.

(b) Geben Sie für den **quadrierten** Parameter σ^2 erwartungstreue Schätzfunktionen $\widetilde{T}_n(X_1, \ldots, X_n)$ an.

(c) Welche Eigenschaft müssen die Schätzfunktionen $\widetilde{T}_n$ aus Teil (b) außerdem erfüllen, um für σ^2 konsistent im quadratischen Mittel zu sein?
(Die Gültigkeit dieser Eigenschaft ist **nicht** *zu überprüfen!)*

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

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

$$f_Y(y|a) = \begin{cases} \dfrac{2(y-1)}{(a-1)^2} & \text{für } 1 \leq y \leq a \\ 0 & \text{sonst} \end{cases}$$

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 $E(Y) = \frac{2}{3}a + \frac{1}{3}$ 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.*
- *Am Ende von Aufgabenteil (b) ist eine Polynomdivision oder eine Erweiterung der rechten Seite eventuell hilfreich.*

Aufgabe 5 (7 + 2 + 3 + 7 = 19 Punkte)

Bei der Herstellung von Scheibenkäse weiß der Hersteller aus Erfahrung, dass die verwendete Produktionsanlage eine Standardabweichung von $5[g]$ für die abgefüllte Menge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Produktionsanlage im Mittel weniger als die auf dem Produkt ausgezeichneten $250[g]$ in die Packungen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 16 Packungen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{16}$ als Realisation einer einfachen Stichprobe vom Umfang 16 zur annahmegemäß $N(\mu, 5^2[g^2])$-verteilten Abfüllmenge betrachtet werden können. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 247.474[g] \; .$$

(a) 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.

(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) Wie groß ist die Wahrscheinlichkeit, bei der Ziehung einer Stichprobe und der anschließenden Durchführung des Tests aus Teil (a) eine Testentscheidung zu Gunsten der Nullhypothese zu erhalten, falls die tatsächliche mittlere Abfüllmenge $246[g]$ beträgt?

(d) Überprüfen Sie unter Verwendung der Varianzschätzung $s^2 = 35.557$ mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$, ob die oben getroffene Annahme $\sigma^2 = 5^2$ aus statistischer Sicht zu verwerfen ist. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

Hinweis: Verwenden Sie für Teil (d) 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
11	3.053	3.816	4.575	10.341	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	11.340	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	12.340	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	13.339	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	14.339	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	15.338	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	16.338	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	17.338	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	18.338	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	19.337	28.412	31.410	34.170	37.566

Aufgabe 6 (14 Punkte)

Um zu überprüfen, ob es einen Zusammenhang zwischen der Impfbereitschaft (eher keine Impfung / eher Impfung) und der Einstellung zu den aktuellen Anti-Corona-Maßnahmen (nicht ausreichend / angemessen / überzogen) gibt, wurden die Ergebnisse einer aktuellen Umfrage, die man als Realisation einer einfachen Stichprobe auffassen können soll, in der folgenden Tabelle zusammengefasst (sämtliche Daten sind **fiktiv**):

	nicht ausreichend	angemessen	überzogen
eher keine Impfung	80	150	120
eher Impfung	120	450	80

Überprüfen Sie anhand dieses Datenmaterials zum Signifikanzniveau $\alpha = 0.05$, ob eine stochastische Abhängigkeit zwischen der Impfbereitschaft und der Einstellung zu den Maßnahmen besteht.

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

Aufgabe 7 (14 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 Sommersemester 2020 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	165	69.25	875241	512.03
2	7	88.64	56749	291.61
3	26	89.88	218891	354.09

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.

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

$n \backslash m$	1	2	3	4	5	195	196	197	198	199
1	161.448	199.500	215.707	224.583	230.162	253.661	253.664	253.667	253.671	253.674
2	18.513	19.000	19.164	19.247	19.296	19.491	19.491	19.491	19.491	19.491
3	10.128	9.552	9.277	9.117	9.013	8.541	8.541	8.540	8.540	8.540
4	7.709	6.944	6.591	6.388	6.256	5.647	5.647	5.646	5.646	5.646
5	6.608	5.786	5.409	5.192	5.050	4.386	4.386	4.385	4.385	4.385
195	3.890	3.042	2.651	2.418	2.260	1.266	1.266	1.266	1.265	1.265
196	3.889	3.042	2.651	2.418	2.260	1.266	1.266	1.265	1.265	1.265
197	3.889	3.042	2.650	2.417	2.260	1.266	1.265	1.265	1.265	1.264
198	3.889	3.042	2.650	2.417	2.260	1.265	1.265	1.264	1.264	1.264
199	3.889	3.041	2.650	2.417	2.259	1.265	1.264	1.264	1.264	1.263

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 2013–2019 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
-24.542  -9.961  18.021  22.432  15.727  -7.730 -13.947

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 157.98483   95.74477    1.65 0.159843
x             1.07205    0.09643   11.12 0.000103 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.16 on 5 degrees of freedom
Multiple R-squared:  0.9611,       Adjusted R-squared:  0.9533
F-statistic: 123.6 on 1 and 5 DF,  p-value: 0.0001026
```

(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 $\alpha = 0.10$, ob β_1 signifikant positiv ist.

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

(f) Welche Warenausfuhrhöhe (in Milliarden Euro) prognostiziert das Modell für ein Jahr mit einer Wareneinfuhrhöhe von 1000 (in Milliarden Euro)?

Aufgabe 9 (6 + 2 + 3 + 3 + 5 = 19 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:

$$\sum_{i=1}^{20} y_i = 318.87; \quad \sum_{i=1}^{20} y_i^2 = 5603.84; \quad \sum_{i=1}^{20} x_i = 97.71;$$

$$\sum_{i=1}^{20} x_i^2 = 524.69; \quad \sum_{i=1}^{20} x_i \cdot y_i = 1710.24$$

(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 symmetrisches Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.90$ für β_1 an.

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

14 Schließende Statistik SS 2021

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

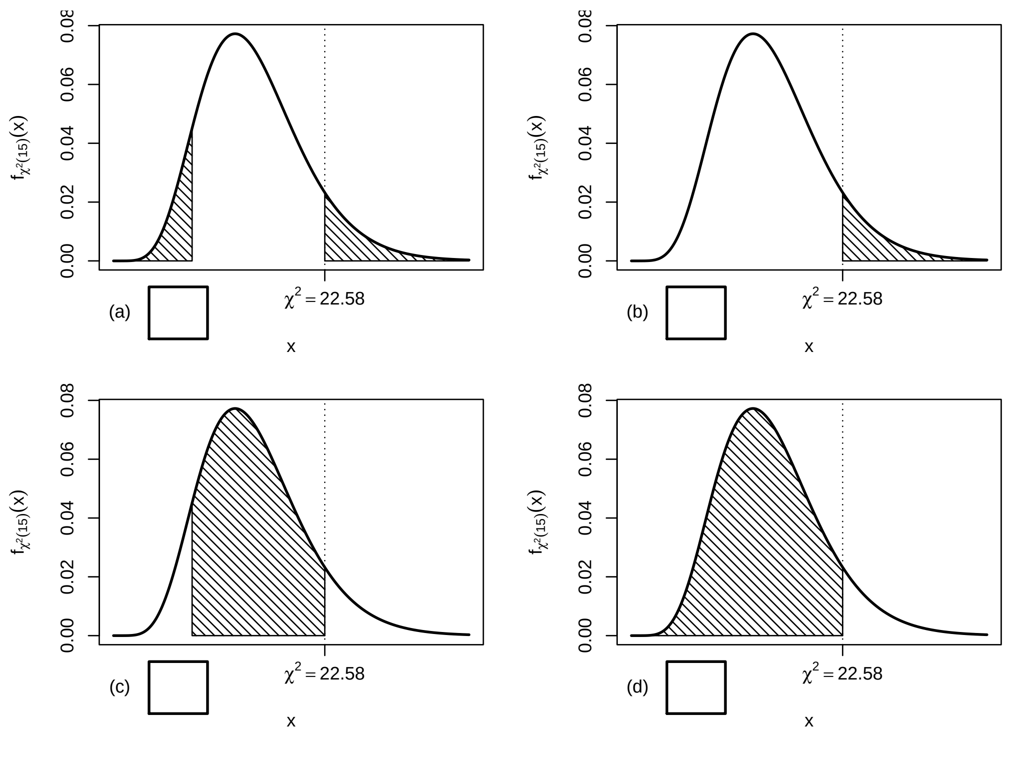
   (c) $\chi^2 \in [\chi^2_{n-1;0.05}, \chi^2_{n-1;0.95}]$ ☐

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

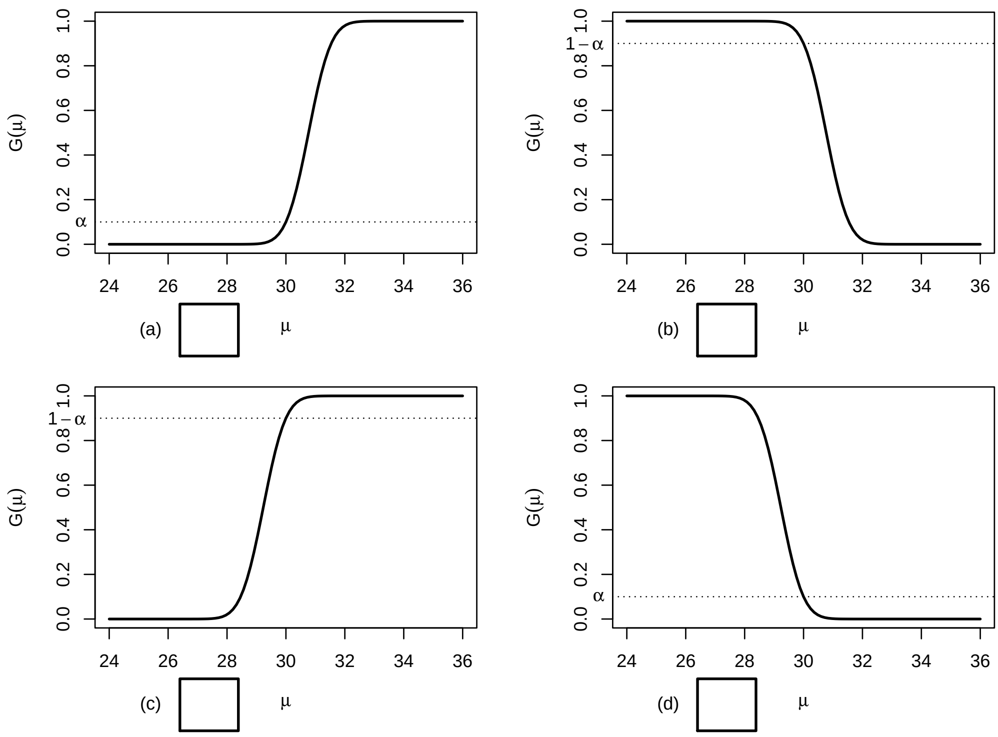
2. Sei $X_1, \ldots, X_{16}$ eine einfache Stichprobe zu einer $N(\mu, \sigma^2)$-verteilten Zufallsvariablen Y mit unbekannten Parametern μ und σ^2. Auf der Grundlage einer Stichprobenrealisation zu dieser einfachen Stichprobe vom Umfang $n = 16$ soll

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

   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 $X_1, \dots, X_{25}$ 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:
Für $n \in \mathbb{N}$ sei $X_1, \dots, X_n$ eine einfache Stichprobe vom Umfang n zu Y.

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

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

(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.

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:

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

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 $E(Y) = \dfrac{2}{3}b + \dfrac{2}{3}$ 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.*

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?

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 $X_1^A, \dots, X_{15}^A$ zu Y^A sowie $X_1^B, \dots, X_{13}^B$ 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.

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 Varianzanalyse 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

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)?

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

Zur Schätzung eines einfachen linearen Regressionsmodells

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

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.

Tabelle zur Standardnormalverteilung

$$F_{N(0,1)}(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$$

$$\Phi(x) = 1 - \Phi(-x)$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

p-Quantile der Standardnormalverteilung

$$F_{N(0,1)}(N_p) = p$$

p	0.85	0.90	0.95	0.975	0.99	0.995	0.9995
N_p	1.036	1.282	1.645	1.960	2.326	2.576	3.291

p-Quantile der $t(n)$-Verteilungen $t_{n;p}$

$$T \sim t(n) \quad \Rightarrow \quad F_T(t_{n;p}) = p$$

$n\backslash p$	0.85	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	1.055	1.310	1.697	2.042	2.457	2.750	3.646
31	1.054	1.309	1.696	2.040	2.453	2.744	3.633
32	1.054	1.309	1.694	2.037	2.449	2.738	3.622
33	1.053	1.308	1.692	2.035	2.445	2.733	3.611
34	1.052	1.307	1.691	2.032	2.441	2.728	3.601
35	1.052	1.306	1.690	2.030	2.438	2.724	3.591
40	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	1.045	1.296	1.671	2.000	2.390	2.660	3.460
80	1.043	1.292	1.664	1.990	2.374	2.639	3.416
100	1.042	1.290	1.660	1.984	2.364	2.626	3.390
120	1.041	1.289	1.658	1.980	2.358	2.617	3.373
140	1.040	1.288	1.656	1.977	2.353	2.611	3.361
160	1.040	1.287	1.654	1.975	2.350	2.607	3.352
180	1.039	1.286	1.653	1.973	2.347	2.603	3.345
200	1.039	1.286	1.653	1.972	2.345	2.601	3.340
250	1.039	1.285	1.651	1.969	2.341	2.596	3.330