DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Zulassungsprüfung Stochastik

der Deutschen Aktuarvereinigung e. V.

am 20. Mai 2023

Hinweise:

- Als Hilfsmittel ist ein Taschenrechner zugelassen.
- Die Gesamtpunktzahl beträgt 120 Punkte. Die Klausur ist bestanden, wenn mindestens 60 Punkte erreicht werden.
- Bitte prüfen Sie die Ihnen vorliegende Prüfungsklausur auf Vollständigkeit. Die Klausur besteht aus 5 Seiten.
- **Alle Antworten sind zu begründen und bei Rechenaufgaben muss der Lösungsweg ersichtlich sein.**

Wir gehen stets von einem Maßraum $(\Omega, \mathcal{A}, \mu)$ bzw. einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ aus. Die Borel σ-Algebra auf $\mathbb{R}^n$ wird mit $\mathcal{B}^n$ bezeichnet, das Lebesgue Maß auf $\mathbb{R}^n$ wird mit λ^n bezeichnet.

Sollten Ihnen in Teilaufgaben Ergebnisse fehlen, dann treffen Sie eine plausible Annahme dafür.

Aufgabe 1. *[24 Punkte]*

Sei $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable und das Maß $\nu : \mathcal{B}^1 \longrightarrow [0,1]$, $\nu(A) = P(X^{-1}(A))$. Beweisen Sie

(a) Für alle $A \subset \mathbb{R}$ gilt

$$1_{X^{-1}(A)} = 1_A \circ X.$$

(b) Für alle Zufallsvariablen $Y : \mathbb{R} \longrightarrow [0, \infty)$ gilt

$$\int_{\mathbb{R}} Y d\nu = \int_{\Omega} Y \circ X dP$$

Aufgabe 2. *[24 Punkte]*

Seien $W_i \sim N(0,1)$, $i = 1,2$ unabhängig, $\rho \in [-1,1]$ und

$$X_1 = W_1$$
$$X_2 = \rho W_1 + \sqrt{1-\rho^2} W_2$$

sowie für $\mu_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1,2$

$$Y_i = \mu_i + \sigma_i X_i, \; i = 1,2.$$

(a) Beweisen Sie

(i) $X_i \sim N(0,1)$, $i = 1,2$

(ii) Für die Korrelation $\rho(X_1, X_2)$ von X_1 und X_2 gilt $\rho(X_1, X_2) = \rho$.

(b) Bestimmen Sie die Verteilung von Y_i, $i = 1,2$ und die Korrelation $\rho(Y_1, Y_2)$ von Y_1 und Y_2.

(c) In der folgenden Teilaufgabe betrachten wir die Zufallsvariablen $I \sim N(100, 15^2)$ bzw. $R \sim N(50, 10^2)$ (Intelligenzquotient bzw. Rechtschreibleistung von Schülern) mit Korrelation $\rho(I, R) = 0,4$. Wir gehen davon aus, dass (I, R) bivariat normalverteilt ist. Sei

$$D = \frac{I - 100}{15} - \frac{R - 50}{10}.$$

(i) Begründen Sie, dass $D \sim N\left(0, \frac{6}{5}\right)$ gilt.

(ii) Aus Untersuchungen ist bekannt, dass ca. 5 % der Schüler an einer Rechtschreibschwäche leiden. Die Diagnose der Rechtschreibschwäche wird gestellt, wenn D einen Schwellwert d_0 überschreitet. Bestimmen Sie d_0 so, dass $P(D > d_0) = 0,05$ gilt.

Aufgabe 3. *[24 Punkte]*

Gegeben seien die Zufallsvariablen X und Θ. Die Dichte von Θ sei

$$f_\Theta(\theta) = \begin{cases} c\theta^k(1-\theta)^l & \theta \in (0,1) \\ 0 & \text{sonst,} \end{cases}$$

wobei $k, l \in \mathbb{N}_0$ und $c > 0$ eine Normierungskonstante ist. Für $\theta \in (0,1)$ sei $X \sim B(n, \theta)$, also ist die bedingte Dichte von X gegeben $\Theta = \theta$ wie folgt:

$$f_{X|\Theta}(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \text{ falls } x \in \{0, \dots, n\}.$$

(a) Bestimmen Sie die gemeinsame Dichte von X und Θ.

(b) Zeigen Sie, dass für die bedingte Dichte $f_{\Theta|X}$ von Θ gegeben $X = x$ gilt:

$$f_{\Theta|X}(\theta|x) = \frac{(n+k+l+1)!}{(k+x)!(n+l-x)!}\theta^{x+k}(1-\theta)^{n+l-x}, \; x \in \{0, \dots, n\}.$$

Hinweis: Für $p, q \in \mathbb{N}_0$ gilt $\int_0^1 t^p(1-t)^q \, dt = \frac{p!q!}{(p+q+1)!}$.

(c) Bestimmen Sie

(i) $E(\Theta|X)$ (ii) $E(\Theta|X = 0)$ für $n = 6$, $k = l = 1$.

Aufgabe 4. *[24 Punkte]*

Sei $\lambda > 0$ und für alle $t > 0$ sei $X_t \sim Exp(\lambda t)$. Gegeben seien $t_1, \dots, t_n > 0$ und $n \geq 2$ unabhängige Beobachtungen X_{t_i}, $i = 1, \dots, n$ wobei $t_i > 0$ gelte.

(a) Zeigen Sie:

(i) $tX_t \sim \mathrm{Exp}(\lambda)$ für alle $t > 0$

(ii) $G := \sum_{i=1}^n t_i X_{t_i} \sim \Gamma(n, \lambda)$

(iii) $E\left(\frac{1}{G}\right) = \frac{\lambda}{n-1}$ mit G aus (ii).

(b) Zeigen Sie, dass $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} t_i X_i}$ ein Maximum-Likelihood Schätzer für λ ist.

(c) Prüfen Sie, ob der Schätzer aus (b) erwartungstreu ist und geben Sie gegebenenfalls eine erwartungstreue Modifikation an.

Aufgabe 5. *[24 Punkte]*

Für den Zeitraum 2018-2021 wird für den (wöchentlichen) Gasverbrauch Z angenommen, dass

$$Z = \alpha_0 + \beta_0 x + \delta \text{ mit } \delta \sim N(0, \rho_0^2) \tag{1}$$

gilt. Hierbei ist x die (durchschnittliche, wöchentliche) Temperatur.

In dieser Aufgabe wird untersucht, ob der Gasverbrauch Y im Zeitraum Oktober-Dezember 2022 im Vergleich zum Gasverbrauch Z der Jahre 2018-2021 um 25 % gesunken ist, also ob

$$Y = 0,75Z \text{ in Verteilung} \tag{2}$$

gilt, d.h. $P(Y \leq y) = P(0,75Z \leq y)$ für alle $y \in \mathbb{R}$.

(a) Beweisen Sie mit (1) und (2), dass

$$Y = a + bx + \varepsilon \text{ mit } \varepsilon \sim N(0, \sigma^2)$$

gilt, und geben Sie a, b, σ^2 in Abhängigkeit von α_0, β_0 und ρ_0^2 an.

In der Folge gehen wir in (1) von folgenden Parametrwerten aus:

$$\alpha_0 = 2513, \qquad \beta_0 = -137, \qquad \rho_0^2 = 33000. \tag{3}$$

(b) Für Oktober bis Dezember 2022 liegen Daten $(x_1, y_1), \ldots, (x_n, y_n)$ mit $n = 15$ vor, siehe auch Abbildung 1, Seite 5. Wir gehen vom Modell der einfachen linearen Regression in (a) aus, also

$$Y_i = a + bx_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2),\ \varepsilon_i \text{ unabhängig},\ i = 1, \ldots, n.$$

Mit einem Statistikprogramm werden die folgenden Schätzwerte bestimmt:

$$\hat{a} = 1954, \qquad \hat{b} = -105, \qquad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 21000,$$

$$\text{se}(\hat{a}) = 65,1, \qquad \text{se}(\hat{b}) = 6,2.$$

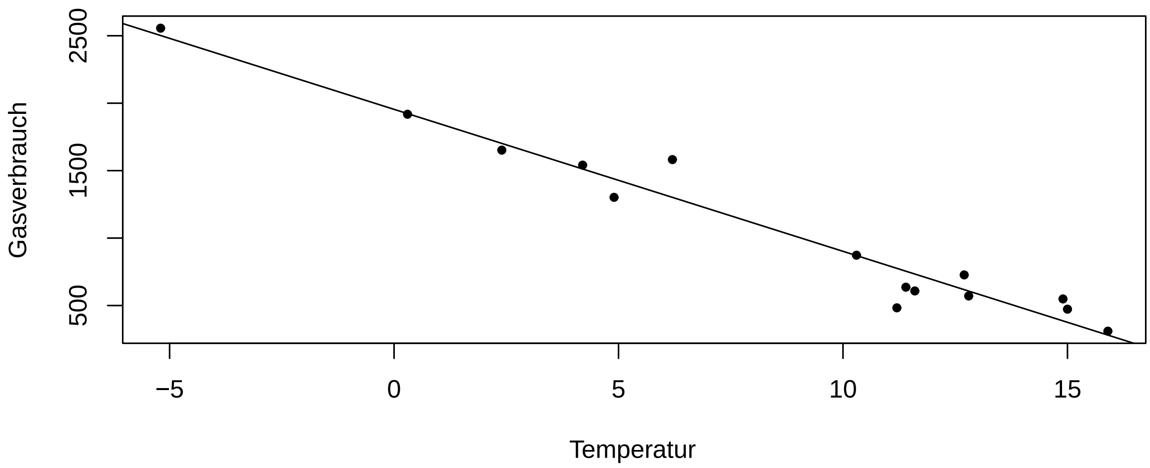

Abbildung 1: Gasverbrauch und Temperatur Oktober-Dezember 2022

(i) Testen Sie die Hypothesen

$$a = 0,75\alpha_0$$
$$b = 0,75\beta_0$$

mit α_0, β_0 aus (3) jeweils zum Niveau von 5 %.

(ii) Sei $S^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$. Es ist bekannt, dass $\frac{S^2}{\sigma^2} \sim \chi^2_{n-2}$ gilt. Beweisen Sie, dass für $p \in (0,1)$

$$P\left(\frac{S^2}{\sigma^2} \in \left[\chi^2_{n-2,p}, \chi^2_{n-2,1-p}\right]\right) = 1 - 2p$$

gilt.

(iii) Testen Sie mit Hilfe von (ii) die Hypothese

$$\sigma^2 = 0,75^2\rho_0^2$$

zum Niveau von 5 %.

(iv) Kommentieren Sie die Ergebnisse von (i) und (iii).

(c) Ein Statistikprogramm gibt mit den Daten aus (b) für die Temperatur von $x = -5,2$ das Konfidenzintervall [2300,2700] zum Niveau 5 % für den erwarteten Gasverbrauch aus. Bei einer Temperatur von $x = -5,2$ wird der Gasverbrauch 2556 beobachtet. Handelt es sich bei diesem Wert um einen unplausiblen Ausreißer?

Lösungsvorschläge

Aufgabe 1 [24 Punkte]

Zu (a) [4 Punkte]

Sei $\omega \in \Omega$ und $A \subset \mathbb{R}$

$$1_{X^{-1}(A)}(\omega) = 1 \iff \omega \in X^{-1}(A) \iff X(\omega) \in A \iff 1_A(X(\omega)) = 1.$$

Zu (b) [20 Punkte]

Sei zunächst Y eine Treppenfunktion, also $Y = \sum_{i=1}^{n} \alpha_i 1_{A_i}$, $\alpha_i \geq 0$, $A_i \in \mathcal{B}^1$. Dann gilt

$$\int_{\mathbb{R}} Y \, d\nu = \sum_{i=1}^{n} \alpha_i \nu(A_i) = \sum_{i=1}^{n} \alpha_i P(X^{-1}(A_i))$$

$$\int_{\Omega} Y(X(\omega)) \, dP(\omega) = \sum_{i=1}^{n} \alpha_i \int_{\Omega} 1_{A_i}(X(\omega)) dP(\omega) \overset{(a)}{=} \sum_{i=1}^{n} \alpha_i \int_{\Omega} 1_{X^{-1}(A_i)}(\omega) \, dP(\omega)$$
$$= \sum_{i=1}^{n} \alpha_i P(X^{-1}(A_i)).$$

Sei nun $Y \geq 0$ messbar. Dann gibt es eine monoton steigende Folge von Treppenfunktionen $Y_n \geq 0$ mit $\lim_{n\to\infty} Y_n = Y$. Dann ist auch $Y_n \circ X$ eine monoton steigende Folge von nicht negativen Funktionen, die gegen $Y \circ X$ konvergiert. Da Y_n Treppenfunktionen sind, gilt

$$\int_{\mathbb{R}} Y_n \, d\nu = \int_{\Omega} Y_n \circ X \, dP \text{ für alle } n \in \mathbb{N}.$$

Laut Satz von der monotonen Konvergenz konvergieren beide Seiten obiger Gleichung

$$\lim_{n\to\infty} \int_{\mathbb{R}} Y_n \, d\nu = \int_{\mathbb{R}} Y \, d\nu$$
$$\lim_{n\to\infty} \int_{\Omega} Y_n \circ X \, dP = \int_{\Omega} Y \circ X \, dP$$

und es folgt die Behauptung.

Aufgabe 2 *[24 Punkte]*

Zu (a) [6 Punkte]

Offensichtlich gilt $X_1 \sim N(0,1)$. Da W_1, W_2 unabhängig sind, sind auch Linearkombinationen von W_1, W_2 normalverteilt, also ist auch X_2 normalverteilt. Ferner gilt

$$\begin{aligned}
E(X_2) &= \rho E(W_1) + \sqrt{1-\rho^2}E(W_2) = \rho 0 + \sqrt{1-\rho^2}0 = 0 \\
Cov(X_1, X_2) &= Cov(W_1, \rho W_1 + \sqrt{1-\rho^2}W_2) = \rho Cov(W_1, W_1) + \sqrt{1-\rho^2}Cov(W_1, W_2) \\
&= \rho Var(W_1) = \rho \\
Var(X_2) &= Var(\rho W_1) + 2Cov(\rho W_1, \sqrt{1-\rho^2}W_2) + Var(\sqrt{1-\rho^2}W_2) \\
&= \rho^2 + 0 + (1-\rho^2) = 1 \\
\rho(X_1, X_2) &= \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} = \rho.
\end{aligned}$$

Zu (b) [8 Punkte]

Wegen der Reproduktivität der Normalverteilung und wegen (a) (i) gilt $Y_i \sim N(\mu_i, \sigma_i^2)$. Damit folgt

$$Y_i - E(Y_i) = \sigma_i X_i \sim N(0, \sigma_i^2),\ i = 1, 2.$$

Wegen $Cov(X_1, X_2) = E(X_1 X_2) = \rho$ gilt

$$Cov(Y_1, Y_2) = E(\sigma_1 X_1 \cdot \sigma_2 X_2) = \sigma_1 \sigma_2 E(X_1 X_2) = \sigma_1 \sigma_2 \rho.$$

Zu (c)

(i) [4 Punkte] Da D eine Linearkombination von I und R ist, und (I, R) bivariat normalverteilt ist, ist auch D normalverteilt. Es handelt sich bei D um die Differenz der Standardisierungen von I und R. Damit folgt

$$\begin{aligned}
E(D) &= 0 \\
Var(D) &= Var\left(\frac{I-100}{15}\right) - 2Cov\left(\frac{I-100}{15}, \frac{R-50}{10}\right) + Var\left(\frac{R-50}{10}\right) = 1 - 2\rho + 1 \\
&= 1,2.
\end{aligned}$$

(ii) [6 Punkte]

$$0,05 = P(D > d_0) = 1 - P(D \leq d_0) = 1 - \Phi\left(\frac{d_0}{\sqrt{1,2}}\right)$$

$$\Phi\left(\frac{d_0}{\sqrt{1,2}}\right) = 0,95 \Longrightarrow \frac{d_0}{\sqrt{1,2}} = u_{0,95} = 1,645 \Longrightarrow d_0 = 1,645\sqrt{1,2} \approx 1,8.$$

Aufgabe 3 [24 Punkte]

Zu (a) [6 Punkte]

Sei $f_\Theta(\vartheta) = c1_{(0,1)}(\vartheta) \cdot \vartheta^k (1-\vartheta)^l$ die Dichte von Θ, f die gemeinsame Dichte von X und Θ. Es gilt für $x \in \{0, \dots, n\}$, $\vartheta \in (0,1)$

$$f(x,\vartheta) = f(x|\vartheta) \cdot f_\Theta(\vartheta)$$

also

$$f(x,\vartheta) = \begin{cases} c\binom{n}{x}\vartheta^{x+k}(1-\vartheta)^{n+l-x} & \text{falls } x \in \{0,\dots,n\}, \vartheta \in (0,1) \\ 0 & \text{sonst.} \end{cases}$$

Zu (b) [8 Punkte]

Die Dichte f_X von X ergibt sich für $x \in \{0, \dots, n\}$ aus

$$\begin{aligned} f_X(x) &= \int_0^1 f(x,\vartheta)\,d\vartheta = c\binom{n}{x}\int_0^1 \vartheta^{x+k}(1-\vartheta)^{n+l-x}\,d\vartheta \\ &= c\binom{n}{x}\frac{(x+k)!(n+l-x)!}{(n+k+l+1)!}. \end{aligned}$$

Damit folgt für $x \in \{0, \dots, n\}$ und $\vartheta \in (0,1)$

$$\begin{aligned} f(\vartheta|x) &= \frac{f(x,\vartheta)}{f_X(x)} = \frac{c\binom{n}{x}\vartheta^{x+k}(1-\vartheta)^{n+l-x}}{c\binom{n}{x}\frac{(x+k)!(n+l-x)!}{(n+k+l+1)!}} \\ &= \frac{(n+k+l+1)!}{(x+k)!(n+l-x)!}\vartheta^{x+k}(1-\vartheta)^{n+l-x}. \end{aligned}$$

Zu (c)

Zu (i) [8 Punkte]

$$\begin{aligned} E(\Theta|X=x) &= \int_0^1 \vartheta f(\vartheta|x)\,d\vartheta \\ &= \frac{(n+k+l+1)!}{(x+k)!(n+l-x)!}\int_0^1 \vartheta^{x+k+1}(1-\vartheta)^{n+l-x}d\vartheta \\ &= \frac{(n+k+l+1)!}{(x+k)!(n+l-x)!}\frac{(x+k+1)!(n+l-x)!}{(n+k+l+2)!} \\ &= \frac{x+k+1}{n+k+l+2}. \end{aligned}$$

Damit ergibt sich $E(\Theta|X) = \dfrac{X+k+1}{n+k+l+2}$.

Zu (ii) [2 Punkte]

Für $k = l = 1$, $n = 6$ und $x = 0$ folgt

$$E(\Theta|X=0) = \frac{2}{10} = \frac{1}{5}.$$

Aufgabe 4 *[24 Punkte]*

Zu (a)

(i) [2 Punkte] Sei $x > 0$. Es gilt

$$P(tX_t \leq x) = P\left(X_t \leq \frac{x}{t}\right) = 1 - e^{-\lambda t \frac{x}{t}} = 1 - e^{-\lambda x}$$

(ii) [2 Punkte] Mit $X_1, \dots, X_n$ sind auch $t_1X_1, \dots, t_nX_n$ unabhängig und aus (i) folgt $t_iX_i \sim Exp(\lambda) = \Gamma(1, \lambda)$. Somit ist die Summe G Gammaverteilt $\Gamma(n, \lambda)$.

(iii) [4 Punkte] Mit (ii) gilt

$$\begin{aligned} E\left(\frac{1}{G}\right) &= \int_0^\infty \frac{1}{x} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}\, dx = \frac{\lambda^n}{\Gamma(n)} \int_0^\infty x^{n-2} e^{-\lambda x}\, dx \\ &= \frac{\lambda^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\lambda^{n-1}} \int_0^\infty \frac{\lambda^{n-1}}{\Gamma(n-1)} x^{n-2} e^{-\lambda x}\, dx = \frac{\lambda}{n-1}. \end{aligned}$$

Die letzte Gleichung gilt, da über die Dichte der $\Gamma(n-1, \lambda)$-Verteilung integriert wird.

(b) [12 Punkte] Es ergibt sich eine Likelihood

$$L(\lambda) = \prod_{i=1}^n \left(t_i \lambda \exp^{-t_i x_i \lambda}\right) = \lambda^n \exp^{-\lambda \sum_{i=1}^n t_i x_i} \prod_{i=1}^n t_i.$$

Somit folgt für $\ell := \ln(L)$

$$\begin{aligned} \ell(\lambda) &= n \ln(\lambda) - \lambda \sum_{i=1}^n t_i x_i + \sum_{i=1}^n t_i \\ \ell'(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^n t_i x_i \\ \ell''(\lambda) &= -\frac{n}{\lambda^2} < 0 \end{aligned}$$

Aus der zweiten Gleichung ergibt sich durch Nullsetzen und Auflösen

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n t_i x_i}$$

und mit $\ell'' < 0$ folgt, dass in $\hat{\lambda}$ ein Maximum vorliegt. Der gesuchte Schätzer ist also

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i X_i}$$

Zu (c) [4 Punkte]

Mit (a) ergibt sich der Erwartungswert von $\hat{\lambda}$

$$E\left(\frac{n}{\sum_{i=1}^{n} t_i X_i}\right) = \frac{\lambda n}{n-1}.$$

Damit ist $\hat{\lambda}$ nicht erwartungstreu. Die Modifikation $\lambda^* = \frac{n-1}{n}\hat{\lambda}$ ist offensichtlich erwartungstreu.

Aufgabe 5 (24 Punkte)

Zu (a) [3 Punkte]

Sei $c = 0,75$. Es gilt für $a = c\alpha_0$, $b = c\beta_0$, $\sigma^2 = c^2\rho_0^2$ und $\varepsilon = c\delta \sim N(0, \sigma^2)$

$$\begin{aligned} P(Y \leq y) &= P(cZ \leq y) = P(c(\alpha_0 + \beta_0 x + \delta) \leq y) = P(c\alpha_0 + c\beta_0 x + c\delta \leq y) \\ &= P(a + bx + \varepsilon \leq y) \end{aligned}$$

und somit die Behauptung.

Zu (b)

(i) [8 Punkte] Es ergeben sich die Testgrößen

$$\frac{\hat{a} - 0,75\alpha_0}{\mathrm{se}(\hat{a})} = \frac{1954 - 0,75 \cdot 2513}{65,1} = 1,06 \qquad \text{(Test für } a\text{)}$$

$$\frac{\hat{b} - 0,75\beta_0}{\mathrm{se}(\hat{b})} = \frac{-105 + 0,75 \cdot 137}{6,2} = -0,36. \qquad \text{(Test für } b\text{)}$$

Die Testgrößen sind t-verteilt mit $n-2 = 13$ Freiheitsgraden, die Hypothesen werden nicht verworfen, da beide Werte im Intervall

$$[-t_{n-2,\frac{975}{1000}}, t_{n-2,\frac{975}{1000}}] = [-2,160,\ 2,160]$$

liegen.

(ii) [3 Punkte]

$$\begin{aligned} P\left(\frac{S^2}{\sigma^2} \in \left[\chi^2_{n-2,p}, \chi^2_{n-2,1-p}\right]\right) &= P\left(\frac{S^2}{\sigma^2} \leq \chi^2_{n-2,1-p}\right) - P\left(\frac{S^2}{\sigma^2} \leq \chi^2_{n-2,p}\right) \\ &= 1 - p - p = 1 - 2p. \end{aligned}$$

(iii) [4 Punkte] Man kann (ii) mit $p = 0,025$ anwenden. Mit der Hypothese ergibt sich die Testgröße

$$\frac{\hat{\sigma}^2(n-2)}{0,75\rho_0^2} = \frac{21000 \cdot 13}{0,75^2 \cdot 33000} = \frac{273000}{18562,5} = 14,7.$$

Sie liegt im Intervall

$$\left[\chi^2_{13,\frac{25}{1000}}, \chi^2_{13,\frac{975}{1000}}\right] = [5,009,\ 24,376]$$

also wird die Hypothese wird nicht verworfen.

(iv) [3 Punkte] Die Daten widersprechen nicht der Annahme, dass der Gasverbrauch um 25 % gesunken ist.

Zu (c) [3 Punkte]

Es handelt sich nicht um einen Aureißer, da 2556 bereits im Konfidenzintervall liegt. Da das Konfidenzintervall eine Teilmenge des Prognoseintervalls ist, ist der Wert von 2556 im der prognostizierten Bandbreite.

Bemerkung: es handelt sich um echte Daten, die die Bundesnetzagentur veröffentlicht hat. Nur die Schätzwerte wurden leicht angepasst um den Rechenaufwand zu verringern.