Probeklausur - Statistik II, SoSe 2018

Aufgabe 1: Mehrdimensionale Zufallsvariablen (15 Punkte)

Gegeben sei ein zweidimensionaler stetiger Zufallsvektor $\mathbf{X} = (X_1, X_2)^T$ mit der gemeinsamen Dichtefunktion

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} x_1 + x_2 & \text{für } 0 \leq x_1 \leq 1 \wedge 0 \leq x_2 \leq 1 \\ 0 & \text{sonst} \end{cases}.$$

a) Bestimmen Sie die Randdichtefunktion $f_{X_1}(x_1)$. (3 Punkte)

b) Berechnen Sie $E[X_1]$. (5 Punkte)

c) Geben Sie $f_{X_2}(x_2)$ und $E[X_2]$ an und begründen Sie Ihre Angaben. (2 Punkte)

d) Berechnen Sie den Erwartungswert $E(X_1 \cdot X_2)$. (5 Punkte)

Aufgabe 2: Punktschätzung (10 Punkte)

Y ist eine stetige Zufallsvariable, die auf dem Intervall $[a, b]$ gleichverteilt ist. Die Dichte von Y ist gegeben durch

$$f_Y(y) := \begin{cases} \frac{1}{b-a} & \text{für } y \in [a, b] \\ 0 & \text{sonst} \end{cases}.$$

Es liegt folgende Stichprobe mit $n = 8$ Beobachtungen vor:

i	1	2	3	4	5	6	7	8
y_i	4	2	5	5	3	4	6	3

Schätzen Sie a und b mittels Momentenmethode.

Aufgabe 3: Ein-Stichproben-Fall (25 Punkte)

Bei einem Test wird die Reaktionszeit eines Kandidaten auf ein Signal durch eine normalverteilte Zufallsvariable X beschrieben. Der Reaktionstest wurde mit diesem Kandidaten sieben Mal durchgeführt. Seine Reaktionszeiten betrugen dabei (in Sekunden):

$$2{,}1 \quad 1{,}8 \quad 2{,}6 \quad 2{,}3 \quad 2{,}2 \quad 2{,}0 \quad 2{,}4$$

a) Schätzen Sie $E(X)$ und $\mathrm{Var}(X)$ nach der Maximum-Likelihood-Methode und geben Sie dabei an, ob es sich um erwartungstreue Schätzer handelt. Geben Sie im Falle verzerrter Schätzungen auch erwartungstreue Schätzer für $E(X)$ und $\mathrm{Var}(X)$ an. (8 Punkte)

b) Die mittlere Reaktionszeit des Kandidaten darf höchstens 2 Sekunden betragen, um im Normbereich zu liegen (H_0). Testen Sie mit einer maximalen Irrtumswahrscheinlichkeit von $\alpha = 0{,}1$, ob die Reaktionszeit des Kandidaten im Normbereich liegt. (8 Punkte)

c) Gehen Sie nun davon aus, dass $\mu = 2{,}1$ und $\sigma^2 = 0{,}07$ gilt, die Reaktionszeit des Kandidaten also nicht im Normbereich liegt. Wie ist $\bar{X}$ dann verteilt? Berechnen Sie zudem die Wahrscheinlichkeit für einen β-Fehler auf Basis des Hypothesenpaares aus Aufgabenteil b) bei einem Ablehnungsbereich von $C := \{(x_1, \ldots, x_7)^T \in \mathbb{R}^7 : \bar{x} \geq 2{,}15\}$. Wie groß ist hierbei die Wahrscheinlichkeit für eine richtige Entscheidung? (9 Punkte)

Aufgabe 4: Test auf Unabhängigkeit (20 Punkte)

Um zu prüfen, ob bei den Besuchern eines Geschäftes die Ausgabenhöhe unabhängig vom Geschlecht ist, wurden zufällig 400 Besucher des Geschäftes ausgewählt und nach Geschlecht und Ausgabenhöhe befragt. 10% bzw. 60% der ausgewählten Besucher gaben höchstens 40 € bzw. höchstens 100 € aus. Die übrigen Besucher gaben über 100 € aus. Von den weiblichen Besuchern gaben 11,25% höchstens 40 € und 43,75% über 100 € aus. 60% der ausgewählten Besucher waren weiblich. Führen Sie den Test mit einer Fehlerwahrscheinlichkeit von $\alpha = 0{,}05$ durch.

Begründen Sie zudem, ob ein restriktiver gewähltes Signifikanzniveau ihre Testentscheidung ändern würde.

Aufgabe 5: Lineare Regression (20 Punkte)

Über 12 Monate hinweg wurden die Werbeausgaben in einem Unternehmen systematisch gesteigert, damit überprüft werden kann, wie der Umsatz (Y) auf die Höhe der Werbeausgaben (X) reagiert (jeweils in €). Aus den Daten wurden bereits folgende Summen berechnet:

$$\sum_{i=1}^{12} x_i = 96 \quad \sum_{i=1}^{12} y_i = 1080 \quad \sum_{i=1}^{12} x_i^2 = 828 \quad \sum_{i=1}^{12} y_i^2 = 123504 \quad \sum_{i=1}^{12} x_i y_i = 9336$$

Unterstellen Sie im Folgenden die lineare Regressionsbeziehung

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

mit $\varepsilon_i \sim N(0, \sigma^2)$ und $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0$ für alle $i, j = 1, \ldots, 12$ und $i \neq j$.

a) Schätzen Sie die Koeffizienten des linearen Regressionsmodells mittels KQ-Methode und interpretieren Sie den Schätzwert für β. (6 Punkte)

b) Testen Sie bei einer maximal zulässigen Wahrscheinlichkeit für einen Fehler 1. Art von 5% die Nullhypothese „Eine Erhöhung der Werbeausgaben um einen Euro steigert den Umsatz um mindestens 15 Euro“ gegen die Alternative, dass der Umsatz um weniger als 15 Euro steigt. Verwenden Sie hierbei die Schätzung $\widehat{\sigma}^2 = 1680$. (9 Punkte)

c) Stellen Sie ein symmetrisches (zentrales) 0,95-Konfidenzintervall für β auf. Verwenden Sie hierbei ebenfalls die Schätzung $\widehat{\sigma}^2 = 1680$. (5 Punkte)