Zusatzübung zur Vorlesung Statistik II
für Studierende der Soziologie und Magisterstudierende

Prof. Dr. Thomas Augustin
Christiane Dargatz

Aufgabe 1

Wir beobachten Jugendliche ab dem Abitur, bis sie von den Eltern in die erste eigene Wohnung ziehen.

(a) Wie könnte in einem Überlebensdauer-Modell eine mögliche Hazardrate aussehen? Begründe Deine Wahl inhaltlich.

(b) Modelliere denselben Sachverhalt durch einen homogenen Markovprozess.

Aufgabe 2

Es soll ein Filter für Spam-Mails getestet werden. Erfahrungsgemäß sind 95% aller Mails Spam. Eine Prüfung hat ergeben, dass 90% der Spam-Mails auch als solche erkannt werden. Leider werden auch 15% der regulären Mails fälschlicherweise als Spam eingestuft.

Berechne die Wahrscheinlichkeit dafür, dass eine als Spam eingestufte Mail tatsächlich Spam ist und interpretiere das Ergebnis!

Aufgabe 3

Eine Zufallsvariable T heißt *weibullverteilt* mit Parametern $\lambda > 0$ und $\alpha > 0$, falls ihre Dichte für $t \geq 0$ folgende Gestalt hat:

$$f(t) = \lambda\alpha(\lambda t)^{\alpha-1} \exp(-(\lambda t)^{\alpha}).$$

Die Survivorfunktion von T lautet $S(t) = \exp(-(\lambda t)^{\alpha})\,, t \geq 0$.

(a) Berechne die Hazardrate von T für $\alpha = 3$!

(b) Skizziere die Hazardrate für $\lambda = 1$! Was gilt allgemein für Hazardraten mit $\alpha > 1$?

Aufgabe 4

Von den Studenten der Universitätsstadt Schlangenhausen weiß man, dass 6% aller Studenten eine Würgeschlange als Haustier haben. Wir ziehen eine Stichprobe von 5 Studenten. Wie groß ist die Wahrscheinlichkeit, dass mindestens 4 Studenten in der Stichprobe eine Würgeschlange als Haustier haben?

Aufgabe 5

Wir wollen die Prüfungsleistungen von Schülern untersuchen, die zusammen das Abitur machen. Liegt hier eine Kopplung von unabhängigen Experimenten vor bzw. was könnte dagegensprechen?

Aufgabe 6

Zur Beurteilung eines Trainingsprogramms zur friedlichen Lösung sozialer Konflikte wird an 5 zufällig herausgegriffenen Personen die jeweilige Gewaltneigung vor und nach der Absolvierung des Programms miteinander verglichen. Zur Messung wird ein normalverteilter Index verwendet, der umso größer ist, je größer die Gewaltneigung ist. Dabei ergaben sich sich folgende Werte:

Person	1	2	3	4	5
vorher	10	15	5	5	20
nachher	0	10	0	5	25

(a) Lässt sich zum Signifikanzniveau von 5% bestätigen, dass die Maßnahme erfolgreich war? (Begründung!)

(b) Welches Ergebnis erhält man, wenn man nur untersuchen will, ob die Maßnahme eine Änderung (nach oben oder nach unten) des Index bewirkt hat?

Aufgabe 7

An 16 zufällig ausgewählten Personen wurde mittels einer geeigneten Skala die Neigung zu autoritärem Verhalten untersucht. (Man darf annehmen, dass es sich hierbei um ein normalverteiltes Merkmal handelt.) Es ergab sich ein Durchschnittswert von 110 Punkten und eine Stichprobenstreuung (Standardabweichung) von 10 Punkten.

(a) Konstruiere zum Sicherheitsniveau von 95% ein Konfidenzintervall für die mittlere Punktezahl.

(b) Jetzt wird zusätzlich bekannt, dass die Stichprobenstreuung gleich der tatsächlichen Streuung ist. Ist dann das in (a) berechnete Konfidenzintervall für die durchschnittliche Punktezahl breiter oder schmaler? (Begründung!)

(c) Wie kann man mit dem Konfidenzintervall aus (a) einfach die Hypothese $H_0 : \mu = 100$ Punkte gegen $H_1 : \mu \neq 100$ Punkte testen?

Aufgabe 8

Bei einer Bank wurde der Zusammenhang zwischen Kreditausfall (Zufallsvariable X) und Geschlecht (Y) untersucht.

	Geschlecht	
	männlich	weiblich
kein Kreditausfall	10300	9700
Kreditausfall	9800	10200

Führe einen geeigneten Test auf Unabhängigkeit der beiden Zufallsvariablen durch (Signifikanzniveau α=0.05)! Interpretiere das Ergebnis auch inhaltlich!

Aufgabe 9

Eine Zufallsvariable X_i heißt *weibullverteilt* mit Parametern $\lambda > 0, \alpha > 0$, falls ihre Dichte folgende Gestalt hat:

$$f(x) = \lambda\alpha(\lambda x)^{\alpha-1}\exp(-(\lambda x)^{\alpha}).$$

Nun betrachtet man die vier Zufallsvariablen X_1, X_2, X_3, X_4 unter der Annahme, dass die Zufallsvariablen unabhängig und identisch verteilt sind. Außerdem wird $\alpha = 1$ festgelegt.

- Erkläre das ML-Schätzprinzip in Worten.
- Finde den Maximum-Likelihood-Schätzwert $\hat{\lambda}$ zu den Realisierungen $x_1 = 4$, $x_2 = 7$, $x_3 = 3$ und $x_4 = 12$!

Aufgabe 10

Erkläre den Hauptsatz der Statistik.

Aufgabe 11

Betrachte den SPSS Output zur linearen Regression über die Nettomiete in Euro (`nm`) von $n = 1750$ Münchner Wohnungen in Abhängigkeit von den Variablen "Wohnfläche" in m^2 (`wfl`) und "kein gekacheltes Bad" (`badkach0`), die als binäre Variable (Ausprägung 0 steht für gekacheltes Bad, Ausprägung 1 für kein gekacheltes Bad) aufgenommen wurde.

Überprüfe, welche Variablen signifikant ($\alpha = 0.05$) zur Erklärung der Nettomiete beitragen, und interpretiere den Output ausführlich.

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		B	Standardfehler	Beta	T	Signifikanz
1	(Konstante)	126,742	58,234		2,176	,033
	wfl	6,900	,734	,745	9,403	,000
	badkach0	-37,316	51,945	-,057	-,718	,475

a. Abhängige Variable: nm