

**142.351 UE**  
**Statistische Methoden der**  
**Datenanalyse**  
**Beispielsammlung**

**R. Frühwirth**

**Institut für Hochenergiephysik**  
**der Österreichischen Akademie der Wissenschaften**  
**A-1050 Wien, Nikolsdorfer Gasse 18**

**Wintersemester 2012/2013**

# Übung 1

## Beispiel 1.1

Eine Münze wird dreimal geworfen. Was ist die Ergebnismenge des Experiments? Welches Ereignis beschreibt die Aussage „Es wird öfter Kopf als Zahl geworfen“?

## Beispiel 1.2

Es seien  $A, B, C$  drei beliebige Ereignisse. Die folgenden Ereignisse sind durch logische Verknüpfung von  $A, B$  und  $C$  darzustellen:

- a)  $A$  tritt nicht ein;
- b) Nur  $A$  tritt ein;
- c) Weder  $B$  noch  $C$  treten ein;
- d) Alle drei Ereignisse treten ein;
- e) Genau eines der drei Ereignisse tritt ein;
- f) Keines der drei Ereignisse tritt ein;
- g) Wenigstens zwei der drei Ereignisse treten ein;
- h) Genau zwei der drei Ereignisse treten ein;
- i) Höchstens zwei der drei Ereignisse treten ein.

## Beispiel 1.3

Es seien  $A$  und  $B$  zwei Ereignisse, die einander weder ausschließen noch implizieren. Wie ist die kleinste Ereignisalgebra beschaffen, die  $A$  und  $B$  enthält?

## Beispiel 1.4

Es seien  $A$  und  $B$  zwei Ereignisse und  $W(\cdot)$  ein Wahrscheinlichkeitsmaß. Beweisen Sie:

- a)  $W(A \cup B) \leq W(A) + W(B)$
- b)  $W(A \cap B) \geq W(A) + W(B) - 1$
- c)  $W(A \cap B') = W(A) - W(A \cap B)$
- d)  $W(A' \cap B') = 1 - W(A) - W(B) + W(A \cap B)$

### Beispiel 1.5

$A$  und  $B$  seien zwei Ereignisse. Zeigen Sie:

- a) Wenn  $A$  und  $B$  unabhängig sind, dann sind auch  $A$  und  $B'$ ,  $A'$  und  $B$ , sowie auch  $A'$  und  $B'$  unabhängig.
- b) Ist  $W(A | B) = W(A | B')$ , so sind  $A$  und  $B$  unabhängig.

### Beispiel 1.6

$A$  und  $B$  seien zwei Ereignisse mit  $W(A) = 3/4$ ,  $W(B) = 2/3$ . Wie groß muß  $W(A \cap B)$  mindestens sein? Wie groß ist  $W(A \cap B)$  unter der Annahme der Unabhängigkeit von  $A$  und  $B$ ? Berechnen Sie für diesen Fall die Wahrscheinlichkeiten der folgenden Ereignisse:

- a) Keines der beiden Ereignisse tritt ein
- b) Genau eines der beiden Ereignisse tritt ein
- c) Beide Ereignisse treten ein
- d) Mindestens eines der beiden Ereignisse tritt ein
- e) Höchstens eines der beiden Ereignisse tritt ein

### Beispiel 1.7 (M)

Laden Sie die Datenmatrix aus der Datei DM1.txt.

- a) Erstellen Sie eine Häufigkeitstabelle, ein Tortendiagramm und ein Stabdiagramm für Merkmal 3.
- b) Erstellen Sie ein Histogramm für Merkmal 2.
- c) Erstellen Sie ein Histogramm und einen Boxplot für Merkmal 4.

### Beispiel 1.8 (M)

Laden Sie den Datensatz aus der Datei DM2.txt.

- a) Berechnen Sie Mittelwert, Median und five point summary.
- b) Berechnen Sie Varianz, Standardabweichung und Interquartilabstand.
- c) Berechnen Sie die Schiefe  $\gamma$  und den Schiefekoeffizienten  $SK$ .
- d) Erstellen Sie ein Histogramm und einen Boxplot der Daten.
- e) Berechnen Sie die empirische Verteilungsfunktion der Daten und stellen Sie sie graphisch dar.
- f) Wie groß ist die Unterschreitungswahrscheinlichkeit des Mittelwerts der Daten?

## Übung 2

### Beispiel 2.1

Beim Bau eines Gerätes werden 5 Widerstände und 4 Kondensatoren verwendet. Die Fehlerwahrscheinlichkeit der Widerstände sei 2%, die der Kondensatoren 3%. Berechnen Sie unter geeigneten Unabhängigkeitsannahmen die Wahrscheinlichkeit, dass mindestens zwei Bauteile fehlerhaft sind.

### Beispiel 2.2

Sie werfen eine symmetrische Münze  $2n$  mal. Wie groß ist die Wahrscheinlichkeit, genau  $n$  mal „Kopf“ zu werfen? Wie verhält sich die Wahrscheinlichkeit für große  $n$ ?

### Beispiel 2.3

Die Wahrscheinlichkeit eines schweren Unfalls im Lauf eines Jahres beträgt bei einer bestimmten technischen Anlage  $1/10000$ . Wie groß ist die Wahrscheinlichkeit dafür, dass beim Betrieb von 30 Anlagen im Lauf von 10 Jahren der Unfall mindestens einmal auftritt?

### Beispiel 2.4

Berechnen Sie die Wahrscheinlichkeit, dass im Lauf von 37 Würfeln mit dem Roulette genau eine 0 auftritt. Wie lautet die Antwort bei einem Roulette mit  $N$  Zahlen? Wie verhält sich die Wahrscheinlichkeit für großes  $N$ ? Wie groß ist die Wahrscheinlichkeit, dass in  $N$  Würfeln jede Zahl genau einmal auftritt?

### Beispiel 2.5

Ein Experiment verwendet eine große Zahl von ICs. Es bezieht diese von drei verschiedenen Herstellern A, B und C, und zwar von A und B je 25%, und von C 50%. Die Wahrscheinlichkeit, dass ein IC mindestens 40000 Stunden fehlerfrei arbeitet, beträgt für die drei Hersteller 0.92, 0.95 und 0.97.

- a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählter IC mindestens 40000 Stunden arbeitet?
- b) Eine IC fällt vor Ablauf der 40000 Stunden aus. Mit welcher Wahrscheinlichkeit stammt er von A, B oder C?

### Beispiel 2.6

In einem Experiment wurde der Trigger so aufgesetzt, dass nur 1% der gewünschten Ereignisse verworfen wird. Die Analyse der aufgezeichneten Ereignisse ergab, dass davon 96% vom erwünschten Typ sind. Insgesamt wurden 92% aller Ereignisse aufgezeichnet.

- a) Wie groß ist der ursprüngliche Anteil an erwünschten Ereignissen?
- b) Wieviel % der Untergrundereignisse wurden vom Trigger verworfen?
- c) Wie groß ist der Anteil an erwünschten Ereignissen an den nicht aufgezeichneten?

### Beispiel 2.7 (M)

Für zwei Merkmale  $A$  und  $B$  gilt:  $W(A) = 0.65$ ,  $W(A|B) = 0.53$ ,  $W(B|A) = 0.34$ . Erzeugen Sie 1000 Vierfeldertafeln mit jeweils 250 Einträgen gemäß diesen Wahrscheinlichkeiten und stellen Sie die Verteilung der Vierfelderkorrelation in einem Histogramm dar. Wie ändert sich die Verteilung, wenn die Anzahl der Einträge verdoppelt wird?

### Beispiel 2.8 (M)

Erläutern Sie den Begriff der Dichte- bzw. der Verteilungsfunktion anhand einer diskreten Zufallsvariablen. Stellen Sie die Dichte- bzw. Verteilungsfunktionen der folgenden Verteilungen graphisch dar:

- a) Binomialverteilung mit  $n = 10, p = 0.3$  sowie  $n = 60, p = 0.2$
- b) Poissonverteilung mit  $\lambda = 3$  sowie  $\lambda = 12$
- c) Geometrische Verteilung mit  $p = 0.25$
- d) Hypergeometrische Verteilung mit  $N = 100, M = 30, n = 10$

Bestimmen Sie Mittelwert und Varianz der Verteilungen.

## Übung 3

### Beispiel 3.1

Lieferungen eines Produktionsbetriebs, bestehend aus Serien zu je 100 Stück, werden vom Empfänger kontrolliert. Es werden Stichproben vom Umfang 5 gezogen, und die Serie wird zurückgewiesen, wenn mindestens ein Stück der Stichprobe mangelhaft ist. Wie groß ist die Wahrscheinlichkeit, dass eine Serie mit 5% Ausschuß zurückgewiesen wird? Rechnen Sie mit und ohne Zurücklegen.

### Beispiel 3.2

Sie entnehmen einer Lieferung von Widerständen zufällig  $n$  Stück, wobei Sie jedes Stück wieder zurücklegen. Sie wissen, dass die Fehlerquote bei der Produktion durch die Zahl  $p$  gegeben ist.

- a) Wie groß ist die Wahrscheinlichkeit, dass Sie  $k$  fehlerhafte Widerstände erwischen?
- b) Wie groß ist die zu erwartende Anzahl von fehlerhaften Stücken, und wie groß ist ihre Varianz?
- c) Wie klein muß  $p$  sein, damit die Wahrscheinlichkeit einer fehlerlosen Stichprobe vom Umfang 10 über 99% steigt?

### Beispiel 3.3

Beim Bridgespiel wird ein Paket von 52 Karten auf 4 Spieler verteilt, sodass jeder Spieler 13 Karten erhält. Wie groß ist die Wahrscheinlichkeit, dass

- a) Sie kein As haben?
- b) Sie mindestens ein As haben?
- c) Sie alle 4 Asse haben?
- d) irgendein Spieler alle 4 Asse hat?
- e) Ihr Partner genau 2 Asse hat?
- f) Sie gemeinsam mit Ihrem Partner alle 4 Asse haben?

### Beispiel 3.4

Sie messen die Aktivität einer Quelle mit einer mittleren Zerfallsrate  $\mu$ . Die Ansprechwahrscheinlichkeit Ihres Detektors beträgt jedoch nur 80%. Wie ist die Zahl der beobachteten Zerfälle verteilt? Wie groß ist ihr Mittelwert? Wie ist die Wartezeit zwischen zwei beobachteten Zerfällen verteilt?

**Beispiel 3.5**

Wie ist die Summe von zwei unabhängigen Poisson-verteilten Zufallsvariablen verteilt? Wie hängt die mittlere Wartezeit des Summenprozesses von den mittleren Wartezeiten der Einzelprozess ab?

**Beispiel 3.6**

Ein Detektor für kosmische Strahlung wird im Mittel von 80 Myonen pro Sekunde getroffen.

- a) Wie ist die Anzahl der Myonen pro Sekunde verteilt?
- b) Wie groß ist die Varianz und die Standardabweichung der Anzahl?
- c) Wie groß ist die Wahrscheinlichkeit, dass in einer Sekunde mehr als 100 Myonen beobachtet werden?
- d) Wie groß ist die Wahrscheinlichkeit, dass in einer Sekunde weniger als 50 Myonen beobachtet werden?

**Beispiel 3.7 (M)**

Erläutern Sie den Begriff der Dichte- bzw. der Verteilungsfunktion anhand einer stetigen Zufallsvariablen. Stellen Sie die Dichte- bzw. Verteilungsfunktionen der folgenden Verteilungen graphisch dar:

- a) Exponentialverteilung mit  $\tau = 2$  und  $\tau = 5$
- b) Normalverteilung mit  $\mu = 0$  und  $\sigma = 1, 3$
- c) t-Verteilung mit  $n$  Freiheitsgraden ( $n = 1, 2, 5, 10$ )

Bestimmen Sie Mittelwert und Varianz der Verteilungen.



## Übung 4

### Beispiel 4.1

Die Lebensdauer eines Bauteils ist exponentialverteilt. Wie groß muß die mittlere Lebensdauer mindestens sein, damit ein Bauteil mit 50% Wahrscheinlichkeit nach einem Jahr noch funktioniert?

### Beispiel 4.2

Die Lebensdauer eines elektrischen Bauteils ist exponentialverteilt mit dem Mittel  $\tau$ . Sie schalten  $N$  gleichartige Bauteile gleichzeitig ein. Wie ist die Wartezeit bis zum ersten Ausfall verteilt? Wie ist die Wartezeit bis zum letzten Ausfall verteilt? Berechnen Sie jeweils Mittelwert und Varianz der Verteilung.

### Beispiel 4.3

Ein Messgerät zur Entfernungsmessung arbeitet mit Präzision  $\sigma^2$ , d.h. der Messwert streut um den wahren Wert  $x$  gemäß einer Normalverteilung  $\mathcal{N}(x, \sigma^2)$ . Wie groß darf  $\sigma$  maximal sein, damit der Messwert vom wahren Wert mit 99% Sicherheit nicht mehr als 1 mm abweicht?

### Beispiel 4.4 (M)

Die Lebensdauer einer Glühbirne (in Stunden) ist normalverteilt mit dem Mittel  $\mu = 1200$  und der Standardabweichung  $\sigma = 25$ . Sie schalten  $N = 25$  gleichartige Birnen gleichzeitig ein. Wie ist die Wartezeit bis zum ersten Ausfall verteilt? Wie ist die Wartezeit bis zum letzten Ausfall verteilt? Überprüfen Sie Ihre Rechnung mittels Simulation und dem Kolmogorov-Smirnov-Test.

### Beispiel 4.5 (M)

Beantworten Sie die Fragen des Beispiels 3.6 mittels Simulation des Experiments, wenn die Aufzeichnung eines Myons eine Totzeit von 10 Millisekunden verursacht, während der keine weiteren Myonen registriert werden können.

### Beispiel 4.6 (M)

Simulieren Sie Ereignisse mit exponentiell verteilten Wartezeiten. Ermitteln Sie die für jede Zeiteinheit die Anzahl  $n$  der Ereignisse und stellen Sie die Häufigkeitsverteilung von  $n$  graphisch dar. Vergleichen Sie mit einer Poissonverteilung. Wie sieht die Verteilung von  $n$  aus, wenn die Wartezeiten gleichverteilt sind?

## Übung 5

### Beispiel 5.1

Sie messen den inversen Impuls  $q = 1/p$  eines Teilchens mit einem relativen Fehler von 10%. Berechnen Sie mit Fehlerfortpflanzung den relativen Fehler des Impulses  $p$ .

### Beispiel 5.2

Sie messen an einem Gleichstrommotor eine Spannung von  $U = 120 \text{ V}$  und eine Stromstärke von  $I = 3.5 \text{ A}$ , jeweils mit einem relativen Fehler von 1%. Berechnen Sie mit Fehlerfortpflanzung den relativen Fehler der Leistung  $P = U \cdot I$ .

### Beispiel 5.3

Der  $n$ -dimensionale Zufallsvektor  $\mathbf{x}$  ist normalverteilt mit Mittel  $\boldsymbol{\mu}$  und Kovarianzmatrix  $\mathbf{V}$ .

- Wie ist eine Komponente  $x_i$  verteilt?
- $\mathbf{H}$  ist eine reguläre  $n \times n$ -Matrix. Wie ist  $\mathbf{y} = \mathbf{H}\mathbf{x}$  verteilt?
- $\mathbf{E}$  ist eine  $m \times n$ -Matrix vom Rang  $m \leq n$ . Wie ist  $\mathbf{z} = \mathbf{E}\mathbf{x}$  verteilt?

### Beispiel 5.4

Der Zufallsvektor  $\mathbf{x}$  ist normalverteilt mit Mittel 0 und Kovarianzmatrix  $\mathbf{V} = \mathbf{G}^{-1}$ . Wie ist  $\mathbf{y} = \mathbf{x}^T \mathbf{G} \mathbf{x}$  verteilt?

### Beispiel 5.5

Erfinden Sie ein Verfahren zur Simulation eines normalverteilten Zufallsvektors mit Mittel  $\boldsymbol{\mu}$  und Kovarianzmatrix  $\mathbf{V}$ .

### Beispiel 5.6 (M)

Die Datei DM3.txt enthält eine normalverteilte Messreihe.

- Schätzen Sie Mittelwert  $\mu$  und Varianz  $\sigma^2$  der Normalverteilung.
- Geben Sie 95%-ige Konfidenzintervalle für beide Größen an.
- Testen Sie die Hypothese  $\mu > 50$  mit dem Signifikanzniveau  $\alpha = 0.01$ .
- Testen Sie die Hypothese  $\sigma < 2$  mit dem Signifikanzniveau  $\alpha = 0.01$ .

### Beispiel 5.7 (M)

In einem Laboratorium wird mit einem Geigerzähler die Hintergrundstrahlung gemessen. Die Datenreihe in der Datei DM4.txt enthält die beobachteten Zählraten (pro Minute).

- a) Schätzen Sie die mittlere Rate pro Minute  $\lambda$  mit der Maximum-Likelihood-Methode.
- b) Geben Sie die Standardabweichung Ihrer Schätzung an.
- c) Testen Sie die Hypothese, dass die mittlere Rate der Hintergrundstrahlung höchstens 60/Minute ist, mit dem Signifikanzniveau  $\alpha = 0.01$ .

## Übung 6

### Beispiel 6.1

Eine bestimmte Operation wurde  $N$  mal ausgeführt und ist dabei  $n$  mal gelungen. Wie würden Sie die Erfolgswahrscheinlichkeit schätzen? Welche Eigenschaften hat die Schätzung?

### Beispiel 6.2

In einer Stichprobe waren 133 von 300 erzeugten Produkte fehlerhaft. Bestimmen Sie das 95% Konfidenzintervall für die Schadenswahrscheinlichkeit

- a) nach Clopper-Pearson und
- b) mit Bootstrapping. Vergleichen Sie auch das robuste Verfahren und die Korrektur von Agresti-Coull.

### Beispiel 6.3

Um die Wahrscheinlichkeit  $p$  eines seltenen Ereignisses zu schätzen, wiederholen Sie einen Versuch so lange, bis das Ereignis eintritt. Schätzen Sie  $p$  aus der Anzahl  $n$  der benötigten Versuche. Welche Eigenschaften hat die Schätzung? Welche Funktion von  $p$  kann erwartungstreu geschätzt werden?

### Beispiel 6.4 (M)

Schätzen Sie den Parameter  $a$  einer Gammaverteilung

$$f(x) = \frac{x^{a-1} e^{-x}}{\Gamma(a)}, \quad x \geq 0$$

mit Hilfe der Maximum-Likelihood-Methode aus einer simulierten Stichprobe vom Umfang 250 ( $a=2$ ). Bestimmen Sie den Fehler der Schätzung durch 2000-malige Wiederholung der Stichprobe. Vergleichen Sie die resultierende Verteilung des Schätzwertes mit einer individuellen Likelihoodfunktion.

### Beispiel 6.5 (M)

Bestimmen Sie empirisch durch Simulation von  $N$  ( $N = 10000$ ) Stichproben vom Umfang  $n$  ( $n = 100$ ) aus einer Standardnormalverteilung die Verteilung des Mittelwerts und des Medians der Stichprobe. Wiederholen Sie die Untersuchung für Stichproben, die mit Ausreißern kontaminiert sind, die einer Normalverteilung mit  $\mu = 0$ ,  $\sigma = 4$  entstammen. Die Wahrscheinlichkeit, dass ein Element der Stichprobe ein Ausreißer ist, soll  $p = 0.2$  sein.

**Beispiel 6.6 (M)**

Bestimmen Sie empirisch durch Simulation von  $N$  ( $N = 10000$ ) Stichproben vom Umfang  $n$  ( $n = 100$ ) aus einer Standardnormalverteilung die Verteilung der Stichprobenvarianz, der Stichprobenstandardabweichung und der Interquartilsdistanz. Wiederholen Sie die Untersuchung für Stichproben, die mit Ausreißern kontaminiert sind, die einer Normalverteilung mit  $\mu = 0$ ,  $\sigma = 4$  entstammen. Die Wahrscheinlichkeit, dass ein Element der Stichprobe ein Ausreißer ist, soll  $p = 0.2$  sein.

## Übung 7

### Beispiel 7.1

Sie wollen die Aktivität (=mittlere Zerfallsrate) einer Quelle bestimmen und messen dazu  $N$  mal die Anzahl der Zerfälle pro Sekunde. Ihr Detektor hat eine Ansprechwahrscheinlichkeit von 75%. Bestimmen Sie die Maximum-Likelihood-Schätzung der Aktivität.

### Beispiel 7.2

Die Stichprobe  $x_1, \dots, x_n$  entstammt einer Gammaverteilung mit bekanntem Formparameter  $a$  und unbekanntem Skalenparameter  $b$ . Berechnen Sie den ML-Schätzer von  $b$ , bestimmen Sie seine Verteilung und berechnen Sie die Fisher-Information der Stichprobe. Welche Eigenschaften hat der Schätzer?.

### Beispiel 7.3

Sie messen die Masse  $m$  einer Probe  $n$  mal und erhalten die Messreihe  $x_1, \dots, x_n$ . Schätzen Sie  $m$  unter der Annahme, dass die Messfehler unabhängig und normalverteilt sind, und zwar

- a) mit der bekannten Varianz  $\sigma^2$ ,
- b) mit unbekannter Varianz,
- c) mit den bekannten Varianzen  $\sigma_1^2, \dots, \sigma_n^2$ . Geben Sie im Fall b) auch eine Schätzung der unbekanntes Varianz an.

### Beispiel 7.4

Die Stichprobe  $x_1, \dots, x_n$  entstammt einer Gleichverteilung im Intervall  $[0, a]$  mit unbekannter oberer Grenze  $a$ .

- a) Berechnen Sie den ML-Schätzer von  $a$  und untersuchen Sie seine Eigenschaften.
- b) Korrigieren Sie den ML-Schätzer so, dass er unverzerrt ist.
- c) Suchen Sie einen unverzerrten Schätzer, der auf dem Stichprobenmittel basiert.
- d) Vergleichen Sie die Varianzen der beiden Schätzer.

**Beispiel 7.5 (M)**

Schätzen Sie die Mittelwerte, die Varianzen und Standardabweichungen, sowie den Korrelationskoeffizienten einer zweidimensionalen Normalverteilung mit Mittel  $\boldsymbol{\mu}=(0,0)$  und Kovarianzmatrix

$$\mathbf{V} = \begin{pmatrix} 1 & 1.2 \\ 1.2 & 4 \end{pmatrix}$$

durch die entsprechenden Stichprobenmomente. Bestimmen Sie durch Simulation von 5000 Stichproben die Verteilung der geschätzten Größen für Stichprobengrößen  $n=25,100,400$ . Wie verhält sich die Verzerrung und die Streuung der geschätzten Größen mit wachsendem  $n$ ?

## Übung 8

### Beispiel 8.1

Das Ergebnis des 1000-maligen Würfeln eines Würfels sind folgende Häufigkeiten: 1: 170, 2: 152, 3: 164, 4: 179, 5: 158, 6: 177. Besteht Grund zu der Annahme, dass der Würfel nicht symmetrisch ist?

### Beispiel 8.2

Sie zählen während eines Jahres die Anzahl der Banküberfälle pro Tag und erhalten die folgende Statistik:

kein Überfall an 310 Tagen  
1 Überfall an 52 Tagen  
2 Überfälle an 3 Tagen  
 $\geq 3$  Überfälle an 0 Tagen

Weicht die Verteilung der Zahl der Überfälle pro Tag signifikant von einer Poissonverteilung ab?

### Beispiel 8.3 (M)

Die folgenden Daten geben die Lebenszeit (in Tagen) von Labormäusen mit Krebstumoren an, die mit einer bestimmten Krebstherapie behandelt wurden.

24	12	36	40	16	10	12	30	38	14	22	18
----	----	----	----	----	----	----	----	----	----	----	----

Entscheiden sie mit Hilfe des Kolmogorov-Smirnow-Tests auf einem Signifikanzniveau von 5% ob die Daten aus einer Log-Normalverteilung mit  $\mu = 3$  und  $\sigma = 4$  stammen.

*Hinweis: Daten sind Log-normalverteilt mit  $\mu$  und  $\sigma$ , wenn ihre natürlichen Logarithmen normalverteilt mit  $\mu$  und  $\sigma$  sind.*

### Beispiel 8.4

Die Zufallsgröße  $s_n$  ist die standardisierte Summe von  $n$  unabhängigen gleichverteilten Zufallsgrößen. Stellen Sie die Dichte von  $s_n$  als Histogramm dar ( $n = 3, 6, 12$ ) und testen Sie, ob die Verteilung von  $s_n$  signifikant von einer Standardnormalverteilung abweicht.



**Beispiel 8.5 (M)**

Schätzen Sie den Parameter  $a$  einer Gammaverteilung

$$f(x) = \frac{x^{a-1} e^{-x}}{\Gamma(a)}, \quad x \geq 0$$

mit Hilfe der gebinnten Methode der kleinsten Fehlerquadrate aus einer simulierten Stichprobe vom Umfang 250 ( $a=2$ ). Bestimmen Sie den Fehler der Schätzung durch 2000-malige Wiederholung der Stichprobe. Vergleichen Sie den so erhaltenen Wert mit dem aus einer individuellen Zielfunktion bestimmten Wert.

## Übung 9

### Beispiel 9.1 (M)

Eine Versicherung interessiert sich für die Abhängigkeit der Schadenswahrscheinlichkeit von der Fahrleistung. Folgende Tabelle zeigt Daten einer Erhebung der jährlichen Fahrleistung  $x$  (in 1000 km) und der Schadenshäufigkeit  $y$  (in %):

$x$	5	10	15	20	25	30	40	50
$y$	9	10	14	18	22	24	29	29

- Stellen Sie die Punkte in einem Streudiagramm dar
- Ermitteln Sie die Regressionsgerade
- Welche Schadenshäufigkeit ist für Fahrer mit 35000 km jährlicher Fahrleistung zu erwarten?

### Beispiel 9.2 (M)

Die folgende Tabelle enthält die Zahl der Sonnenflecken, die zwischen 1970 und 1983 festgestellt wurden sowie die Anzahl der Verkehrstoten durch Autounfälle in diesen Jahren in den USA in tausend. Testen Sie die Hypothese, dass die Zahl der Verkehrstoten nicht von der Zahl der Sonnenflecken beeinflusst wird. Interpretieren Sie Ihr Ergebnis.

Jahr	Sonnenflecken	Verkehrstote
70	165	54.6
71	89	53.3
72	55	56.3
73	34	49.6
74	9	47.1
75	30	45.9
76	59	48.5
77	83	50.1
78	109	52.4
79	127	52.5
80	153	53.2
81	112	51.4
82	80	46.0
83	45	44.6

(Quelle: Sheldon M. Ross, *Statistik f. Ingenieure und Naturwissenschaftler*)

### Beispiel 9.3 (M)

Die Lebensdauer einer Maschinenkomponente hängt von der Betriebsspannung  $x_1$ , der Motordrehzahl  $x_2$  und der Betriebstemperatur  $x_3$  ab. Ein Experiment ergibt die folgende Lebensdauer in Stunden:

$y$	$x_1$	$x_2$	$x_3$
2145	110	750	140
2155	110	850	180
2220	110	1000	140
2225	110	1100	180
2260	120	750	140
2266	120	850	180
2334	120	1000	140
2340	130	1000	180
2212	115	840	150
2180	115	880	150

- Passen Sie ein lineares Modell an die Daten an.
- Bestimmen Sie die Kovarianzmatrix der Regressionsparameter.
- Welche Regressionsparameter sind signifikant von Null verschieden?
- Bestimmen Sie ein 95%-iges Konfidenzintervall für die erwartete Lebensdauer, wenn  $x_1 = 125$ ,  $x_2 = 900$  und  $x_3 = 160$ .

(Aus: Sheldon M. Ross, Statistik für Ingenieure und Naturwissenschaftler)

### Beispiel 9.4 (M)

Die folgende Tabelle enthält das Körpergewicht und das Gewicht des Gehirns von 28 Tieren. Es soll untersucht werden ob man für einen größeren Körper ein größeres Gehirn braucht.

Spezies	Körpergewicht [ $kg$ ]	Gewicht des Gehirns [ $g$ ]
Stummelschwanzhörnchen	1.35	8.1
Kuh	465	423
Grauwolf	36.33	119.5
Ziege	27.66	115
Meerschweinchen	1.04	5.5
Diplodocus	11700	50
Asiatischer Elefant	2547	4603
Esel	187	419
Pferd	521	655
Gibraltar-Affe	10	115
Katze	3.3	25.6
Giraffe	529	680
Gorilla	207	406
Mensch	62	1320
Afrikanischer Elefant	6654	5712
Triceratops	9400	70
Rhesusaffe	6.8	179
Känguruh	35	56
Hamster	0.12	1
Maus	0.023	0.4
Hase	2.5	12.1
Schaf	55.5	175
Jaguar	100	157
Schimpanse	52.16	440
Brachiosaurus	87000	154.5
Ratte	0.28	1.9
Maulwurf	0.122	3.0
Schwein	192	180

Stellen Sie die Logarithmen der Gewichte grafisch dar und bestimmen Sie die Regressionsgerade mit LS und LMS. *Hinweis:* Benützen Sie die MATLAB-Funtion `make_robust_regression`.