Technische Universität Dortmund
Fakultät Statistik
Lehrstuhl Computergestützte Statistik

Dr. Uwe Ligges
M.Sc. Daniel Horn
M.Sc. Hendrik van der Wurp
Steffen Maletz

Übung zur Vorlesung

Computergestützte Statistik

Wintersemester 2018/2019

Übungsblatt Nr. 4

Abgabe ist Montag der 05.11.2018 an CS-abgabe@statistik.tu-dortmund.de oder Briefkasten 138

Aufgabe 1 **(4 Punkte)**

Wie Sie in der Vorlesung gelernt haben, gelten viele mathematische Gesetze, die für Sie eigentlich selbstverständlich sind, in `R` nicht. So haben Sie in der Vorlesung zum Beispiel gesehen, dass $0.1 + 0.2 \neq 0.3$ ist. Finden Sie weitere solche Zahlenbeispiele für die folgenden Situationen (die Beispiele aus dem Skript geben natürlich keine Punkte).

i) (0.5 Punkte) $a + b \neq c$ mit $a, b, c \in \mathbb{R}$ und $a + b = c$ gilt analytisch
(0.1 + 0.2 = 0.3 gibt keine Punkte, das Beispiel hatten wir ja schon),

ii) (0.5 Punkte) $a\frac{b}{b} \neq a$ mit $a, b \in \mathbb{R}$,

iii) (0.5 Punkte) $\sqrt{x}^2 \neq x$ mit $x \in \mathbb{R}$,

iv) (0.5 Punkte) $a + b - b \neq a$ mit $a, b \in \mathbb{R}$,

v) (0.5 Punkte) $exp(log(x)) \neq x$ mit $x \in \mathbb{R}$,

vi) (0.5 Punkte) $\binom{n}{k} \neq \frac{n!}{k!(n-k)!}$ mit $n, k \in \mathbb{N}$.

vii) (1 Punkt) Überlegen Sie sich 2 weitere derartige Beispiele, in denen `R` nicht exakt rechnet.

Aufgabe 2 **(4 Punkte)**

Betrachten Sie quadratische Funktionen der Art $g(x, a) = a_2x^2 + a_1x + a_0$. Seien weiterhin x^* und a_i^* die entsprechenden Repräsentation der Zahlen im Computer mit relativen Fehlern δ_x bzw. δ_{a_i}. Schätzen Sie den relativen Fehler ab, der bei der Auswertung der Funktion gemacht wird. Für welche Werte von a, x können große Fehler auftreten? Geben Sie ein passendes Beispiel in `R` an.

Aufgabe 3 **(4 Punkte)**

Wir vergleichen die beiden Standard-Algorithmen zur Varianzberechnung. Die beiden Implementierung müssen jeweils wie üblich dokumentiert werden. Ein Test ist hier nicht notwendig, die nachfolgende Simulation kann als Test angesehen werden. Neben Implementierung der beiden Algorithmen brauchen wir dafür Vektoren, deren exakte Varianz wir kennen. Dazu verwenden wir Vektoren der Art $[1, 2, ..., n]$.

a) (0.5 Punkte) Implementieren Sie des Two-Pass Algorithmus zur Varianzberechnung: $s(x) = \sum(x_i - \bar{x})^2$.

b) (0.5 Punkte) Implementieren Sie den Textbook Algorithmus zur Varianzberechnung: $s(x) = \sum x_i^2 - \frac{1}{n}(\sum x_i)^2$.

c) (1 Punkt) Berechnen Sie (analytisch) die Varianz des Vektors $[1, ..., n]$ mit n beliebig.

d) Berechnen Sie für je 10 zufällige Permutationen des Vektors $[1, ..., n]$ sowohl mit Ihren Implementierungen aus a) und b) sowie mit der `var`-Funktion die Varianz. Bestimmen Sie die absolute Differenz von der wahren Varianz, die Sie mit der Formel aus c) ausrechnen können Stellen Sie die Fehler für $n \in \{2^1, 2^2, ..., 2^{26}\}$ grafisch dar. Bis wann liefern die Algorithmen exakte Ergebnisse? Welchen Algorithmus implementiert `R` vermutlich? Welchen Einfluss hat die Reihenfolge des Vektors?

Hinweis: Sie dürfen gerne die Funktionen `sum` und `mean` verwenden.