

Online-Test 1: Einleitung, R-Tutorial und statistische Grundlagen

Test und Assessment – Druckansicht

Online-Test 1: Einleitung, R-Tutorial und statistische Grundlagen

Datum: Tue Nov 2 07:52:41 2021 Maximale Punktezahl: 10

Frage 1 - Supervised vs Unsupervised (1 Punkt) [ID: 924948]

Ordnen Sie die verschiedenen Data-Mining-Ansätze den Paradigmen supervised und unsupervised zu!

- | | | | |
|--------------|----------|-------------------------|--------------|
| supervised | passt zu | Klassifikation | (0.2 Punkte) |
| supervised | passt zu | Regression | (0.2 Punkte) |
| unsupervised | passt zu | Association Rule Mining | (0.2 Punkte) |
| unsupervised | passt zu | Clustering | (0.2 Punkte) |
| unsupervised | passt zu | Outlier Detection | (0.2 Punkte) |

Bei Klassifikation und Regression gibt es eine Zielvariable (kategorisch bzw. numerisch), bei den anderen Ansätzen nicht.

Frage 2 - Aggregationsfunktionen (1 Punkt) [ID: 924951]

Welche der folgenden Aggregationsfunktionen sind distributiv, d.h. man kann ihre Ergebnisse auf Teilmengen berechnen und aus den Teilergebnissen dann das Gesamtergebnis erlangen, ohne dass weitere Informationen gespeichert werden müssen?

- ☒ Minimum (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)
- ☐ Mittelwert (Durchschnitt) (*Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.2 Punkte*)
- ☐ Median (*Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.2 Punkte*)
- ☐ Modalwert (*Ausgewählt = 0 Punkte, Nicht ausgewählt = 0.2 Punkte*)
- ☒ Maximum (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)

min()-Teilergebnisse lassen sich durch min() aggregieren, max()-Teilergebnisse durch max().

Für den Mittelwert muss neben den Mittelwerten pro Teilmenge noch ein weiteres Aggregat, z.B. Anzahl der Elemente in der Teilmenge, gespeichert werden (für den Fall, dass die Teilmengen nicht gleich groß sind).

Für Modalwert und Median benötigt man die Anzahl der verschiedenen Werte pro Teilmenge, d.h. im schlimmsten Fall muss trotzdem jeder Wert einzeln gespeichert werden.

Frage 3 - Boxplot (1 Punkt) [ID: 924953]

Welche Quantile sind aus einem Boxplot ablesbar?

- ☒ 0% (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)
- ☒ 25% (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)
- ☒ 50% (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)
- ☒ 75% (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)
- ☒ 100% (*Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte*)

0%-Quantil und 100%-Quantil entsprechen Minimum und Maximum.

Frage 4 - Korrelation und Unabhängigkeit (1 Punkt) [ID: 924955]

Welcher der folgenden Zusammenhänge gilt zwischen zwei Zufallsvariablen?

- ☐ nicht korreliert \iff unabhängig

(0 Punkte)

☐ nicht korreliert \implies unabhängig

(0 Punkte)

☒ unabhängig \implies nicht korreliert

(1 Punkt)

Korrelation ist ein Spezialfall der Abhängigkeit.

Frage 5 - Nullhypothesen bei statistischen Tests (1 Punkt) [ID: 925031]

Ordnen Sie die Nullhypothesen den statistischen Tests (entsprechend der Definition aus der Vorlesung) zu!

Die Zufallsvariablen sind unabhängig. passt zu Chi-Quadrat-Test (0.34 Punkte)

Die Verteilungen unterscheiden sich nicht. passt zu Kolmogorov-Smirnov-Test (0.33 Punkte)

Die Mediane der Verteilungen unterscheiden sich nicht. passt zu Wilcoxon-Mann-Whitney Test (0.33 Punkte)

Siehe Vorlesung.

Frage 6 - length() (1 Punkt) [ID: 924967]

Was ist das Ergebnis von **length(2)** in R?

☐ eine Exception (0 Punkte)

☐ 0 (0 Punkte)

☒ 1 (1 Punkt)

☐ 2 (0 Punkte)

☐ 3 (0 Punkte)

Da es in R keine echten elementaren Werte gibt, ist **2** ein Vektor der Länge 1.

Frage 7 - Boolean-Konstanten (1 Punkt) [ID: 925027]

Wie erhalte ich die Konstante mit dem Wahrheitswert "true" in R?

☐ true (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0 Punkte)

☐ True (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0 Punkte)

☒ TRUE (Ausgewählt = 0.5 Punkte, Nicht ausgewählt = 0 Punkte)

☒ T (Ausgewählt = 0.5 Punkte, Nicht ausgewählt = 0 Punkte)

Die Herausforderung ist nur, Java, Python, R etc. nicht zu verwechseln ;-)

Frage 8 - Typcasting (1 Punkt) [ID: 924999]

Was ist das Ergebnis von **class(c(0, 1, TRUE))** in R?

☐ logical (0 Punkte)

☐ integer (0 Punkte)

☒ numeric (1 Punkt)

☐ character (0 Punkte)

An sich wird beim automatischen Typcasting der spezifischste Typ gewählt, und zwar **logical** vor **integer** vor **numeric** vor **character**. Ganzzahlige numerische Konstanten (nicht aber Sequenzen wie **1:5**) sind aber **numeric**, es sei denn, man wendet **as.integer()** an oder hängt ein **L** an.

Frage 9 - Indizierung (1 Punkt) [ID: 925004]

Was ist das Ergebnis von **letters[-2]** in R?

☐ x (0 Punkte)

☐ y (0 Punkte)

☒ Alle Kleinbuchstaben ohne b. (1 Punkt)

- ☐ Alle Kleinbuchstaben ohne c. (0 Punkte)
- ☐ Alle Kleinbuchstaben ohne x. (0 Punkte)
- ☐ Alle Kleinbuchstaben ohne y. (0 Punkte)

Falls Python-Vorkenntnisse vorhanden sind, kann die negative Indizierung in R verwirrend sein. Es wird nicht von hinten gezählt, sondern weiter von vorne, aber das genannte Element / die genannten Elemente werden entfernt. Außerdem beachten, dass die Indizes mit 1 anfangen.

Frage 10 - Zugriff auf Listenelement (1 Punkt) [ID: 925025]

Gegeben sei die Variable **x = list(alpha=1, beta=2, gamma=3)** in R. Welche der folgenden Zugriffe sind prinzipiell zulässig, d.h. lösen zumindest keine Exception aus?

- ☒ x[2] (Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte)
- ☒ x[[2]] (Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte)
- ☒ x[-2] (Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte)
- ☐ x[[-2]] (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0 Punkte)
- ☐ x[1000] (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0 Punkte)
- ☒ x['alpha'] (Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte)
- ☒ x\$g (Ausgewählt = 0.2 Punkte, Nicht ausgewählt = 0 Punkte)
- ☐ x[:] (Ausgewählt = 0 Punkte, Nicht ausgewählt = 0 Punkte)

Zugriff ist mit einfachen Klammern auf Elemente und mit doppelten Klammern auf Teillisten möglich. Es können positive/negative Indizes verwendet werden, jedoch mit Exception oder "merkwürdigem" Ergebnis wenn außerhalb der eigentlichen Indizes zugegriffen wird. Außerdem können die Namen der Listenelemente verwendet werden, sogar noch mit einer zusätzlichen Notation, der \$-Notation. Abkürzungen von Namen sind in der \$-Notation erlaubt (wenn auch nicht zu empfehlen).