

Statistik und Wahrscheinlichkeitsrechnung (WS 2021/22)

Aufgabenblatt 1

zu bearbeiten bis: 31.10.2021, 23:59 Uhr

Generelle Hinweise I: Abgaben

Bearbeiten Sie die Übungsblätter alleine, oder gerne auch in Zweierteams. Laden Sie Ihre Datei(en) bis Sonntag abend um 23:59 Uhr in den passenden Abgabe-Ordner ("Abgabe Blatt X") in Stud.IP hoch. Fassen Sie alle Theorie-Aufgaben eines Blatts in einer einzigen, max. 10 MB großen, PDF- oder JPEG-Datei zusammen. Name der Datei:

Vorname.Nachname.pdf bzw.

Vorname1.Nachname1-Vorname2.Nachname2.pdf

Zum Beispiel: Angela.Merkel-Lothar.Matthäus.pdf. Beinhaltet das Blatt ein Python-Notebook (siehe unten), so benennen Sie dieses analog:

Vorname.Nachname.ipynb bzw.

Vorname1.Nachname1-Vorname2.Nachname2.ipynb

Generelle Hinweise II: Python-Notebooks

Zur Bearbeitung von Python-Aufgaben nutzen wir sogenannte Python-Notebooks, in denen man Aufgabenstellung und Python-Code wiki-artig mischen kann. Ich werde die Notebooks über die Data Science - Plattform [kaggle.com](https://www.kaggle.com) veröffentlichen, so dass Sie sie direkt im Browser bearbeiten können (Sie können die Notebooks aber auch herunterladen und lokal bearbeiten, z.B. mit der Software miniconda). Sie erhalten jeweils einen Link zu einem Notebook, z.B.

<https://www.kaggle.com/aulges/statwr-00-hello-world>

Klicken Sie auf "Copy&Edit" rechts oben. Es öffnet sich ein Editor, in dem Sie eine Kopie des Notebooks bearbeiten können und so die Aufgabe lösen können. Anschließend laden Sie unter "File > Download Notebook" das fertige Notebook im .ipynb-Format herunter und geben es in Stud.IP ab.

Hinweis: Sollten Sie die Notebooks zur späteren Wiederverwendung auf Kaggle speichern wollen, müssen Sie sich einen Account erstellen und sich einloggen.

Bearbeiten Sie nun das kurze Tutorial-Notebook "Hello World" (siehe Link oben).

Aufgabe 1.1 (Mini-Stichprobe)

Gesucht ist eine sortierte Stichprobe mit $n = 5$ Samples. Das 10%-Quantil beträgt 10. Das 40%-Quantil beträgt 13, der Median 16, die Spannweite 10, der Mittelwert 15. Leiten Sie sämtliche Werte der Stichprobe her.

Aufgabe 1.2 (Notebook)

Bearbeiten Sie das Notebook:

<https://www.kaggle.com/aulges/statwr-01-cars-on-e-bay>

Aufgabe 1.3 (Transformationen)

Gegeben sei eine Stichprobe $x_1, \dots, x_n \in \mathbb{R}$ mit Mittelwert 10 und Varianz 25. Schlagen Sie eine lineare Transformation $x' = \alpha \cdot x + \beta$ vor, so dass der Mittelwert \bar{x}' der transformierten Stichprobe gleich 0 und ihre Varianz $s'^2 = 1$ ist.

Aufgabe 1.4 (Mittelwerte verteilt berechnen)

Wir wollen den Mittelwert von $n = 100.000.000$ (= 100 Millionen) Zahlen berechnen. Die Werte sind über $m = 1000$ Worker-Maschinen verteilt. Jede Worker-Maschine i kann auf einen Teil $x_1^i, x_2^i, \dots, x_{n_i}^i$ der Daten (bestehend aus n_i Zahlen) zugreifen. Es gilt: $n_1 + n_2 + \dots + n_m = n$.

Jeder Worker i berechnet auf "seinen" Daten den Mittelwert \bar{x}_i . Geben Sie eine Formel an, um den Gesamt-Mittelwert aller Zahlen \bar{x} aus den einzelnen Mittelwerten der Worker ($\bar{x}_1, \dots, \bar{x}_m$) und aus n_1, \dots, n_m zu berechnen. Zeigen Sie außerdem die Korrektheit Ihrer Formel.