

Mining von Softwarearchiven mit BOA

In dieser Übung sollen Sie analysieren, ob es Unterschiede bzgl. der Verwendung von TRY-Statements zwischen Java- und Python-Projekten in GitHub gibt. Bitte geben Sie nur die Skripte und Ihren Bericht ab, nicht die berechneten Datensätze!

Aufgabe 1: Relative Häufigkeit mit BOA (4 Punkte)

Implementieren Sie ein BOA-Skript, das für jedes Projekt in temporären Variablen (vom Typ float) die Anzahl der Statements und Try-Statements für den aktuellen Stand jedes Projektes (also alle letzten Dateirevisionen) berechnet und am Ende das Verhältnis dieser beiden Zahlen zurückliefert. Anschließend führen Sie Ihr Skript auf den Datensätzen 2022 Feb/Python und 2019 October/GitHub aus und laden die Ausgabe herunter.

Aufgabe 2: Statistische Auswertung mit Python (3 Punkte)

Für diese Aufgabe erweitern Sie das Python-Skript `uebung6.py`, das Sie auf der Vorlesungswebseite finden. Da wir zwei sehr große Datensätze (aus Aufgabe 1) haben, gehen wir von der schwächeren **Annahme aus, dass diese nicht normalverteilt** ist und verzichten darauf auf Normalverteilung zu testen. Als Signifikanzniveau verwenden Sie $\alpha = 1\%$.

- a) Wählen Sie einen geeigneten statistischen Test mit Hilfe des Entscheidungsbaums auf Folie 29 des Foliensatzes zur quantitativen Analyse aus und berechnen Sie mit der in den Modulen `scipy.stats` oder `researchpy` verfügbaren Implementierung dieses Tests, ob der Unterschied der Mittelwerte für Java- und Python-Projekte statistisch signifikant ist. Der Test liefert Ihnen auch die Effektstärke (Cohen's d).
- b) Bestimmen Sie experimentell eine Schranke für die Sample-Größe (Parameter `sampleSize` von `readSampe()`), ab der der statistische Test nicht mehr signifikant ist.

Aufgabe 3: Interpretation (1 Punkte)

Interpretieren Sie die Effektstärke für die gesamte Stichprobe (Aufgabe 2a). Welchen Schluss kann man aus Aufgabenteil 2b für Studien zu ähnlichen Fragestellungen ziehen.

Abgabe: Donnerstag, der 26.01.2023 zum Vorlesungsbeginn via StudIP

Abgabeformat:

Bei der Abgabe Ihrer Bearbeitung der Übungsaufgaben und Projekte müssen Sie sich an die unter <https://www.st.uni-trier.de/lectures/generic/misc/submission-guidelines.pdf> einsehbare Regelung halten.