



Übung zu Digitalisierung und Statistik

Abgabe bis: Donnerstag, 30. Jänner 2025 (einschließlich)

Aufgabe 14.1 (10 Punkte)

Wir sind Mitarbeiter eines bekannten Pharma-Konzerns in Österreich dessen Namen Rovartis wir aus Verschwiegenheitsgründen hier nicht erwähnen dürfen. Ein neues und vielversprechendes Medikament zur Behandlung von Flatus incarceratus (FI) ist in der Entwicklung. In einer sehr sehr groß angelegten Studie wollen wir den Effekt des Medikaments auf die Zusammensetzung und Konzentration von Proteinen untersuchen. Hierzu haben wir die Konzentration von über 3000 Proteinen an 3 FI-Patienten ohne Einnahme des Medikaments (*control*) und 4 FI-Patienten die das Medikament eingenommen haben (*treated*) gemessen.

1. (3 Punkte) Laden Sie hierzu die Tabelle „Proteomics-Experiment.xlsx“ ein. Löschen Sie alle Zeilen, wo in allen Zellen der Behandelten die Werte fehlen und danach die Zeilen, wo in allen Zellen der Unbehandelten die Werte fehlen. Achtung die neu entstandene Tabelle wird weiterhin NaN Werte enthalten.
2. (1 Punkt) Teilen Sie die bereinigte Tabelle in zwei kleine Tabellen, eine nur mit den unbehandelten und eine nur mit den behandelten Patienten. Transponieren Sie beide Tabellen.
3. (2 Punkte) Führen Sie mit beiden Tabellen zusammen einen multiplen T-Test durch. Nutzen Sie hierzu die richtige `nan_policy`. Schreiben Sie in die Konsole, wie viele Proteine statistisch signifikant sind, also unterhalb $\alpha = 0.05$ liegen.
4. (1 Punkt) Erstellen Sie ein Histogramm der pValue Verteilung.
5. (2 Punkte) Führen Sie eine pValue Korrektur mittels Benjamini-Hochberg durch. Schreiben Sie erneut in die Konsole, wie viele Proteine nach der Korrektur statistisch signifikant sind (Kontrolle: weniger als 10).
6. (1 Punkt) Erstellen Sie erneut ein Histogramm der korrigierten pValues.