



Übung zu Digitalisierung und Statistik

Abgabe & Bearbeitung: Mittwoch, 22. Jänner

Aufgabe 13.1 (7 Punkte)

Gegeben sei die Excel-Datei „Testdaten-Mittwoch.xlsx“. Wir wollen den Effekt eines Cholesterol-senkenden Mittels auf Personen mit normalen und erhöhten Cholesterolwerten überprüfen. Dazu wurde bei 32 Personen mit normalem Cholesterolspiegel zum Zeitpunkt T0 (direkt vor der Einnahme) und bei T1 (1 Tag nach Einnahme) deren Cholesterolkonzentration (in mg/dl) im Blut gemessen. Das Gleiche wurde bei 17 Personen mit erhöhtem Cholesterolspiegel durchgeführt. Lesen Sie das Datenblatt „Cholesterol“ ein.

1. (2 Punkte) Plotten Sie mit Matplotlib einen Plot mit den Histogrammen beider Kohorten zum Zeitpunkt T0 und einen weiteren Plot zum Zeitpunkt T1. Machen Sie die Histogrammbalken wieder halbtransparent.
2. (1 Punkt) Erstellen Sie eine Funktion mit dem Funktionsnamen `signifikanz`, die als Eingabe eine Zahl hat. Wenn die Zahl ≤ 0.001 ist, dann soll der String 'hoch signifikant' zurückgegeben werden. Ansonsten wenn die Zahl ≤ 0.01 ist, dann soll der String 'sehr signifikant' zurückgegeben werden. Ansonsten wenn die Zahl ≤ 0.05 ist, dann soll der String 'signifikant' zurückgegeben werden. Ansonsten soll der String 'nicht signifikant' zurückgegeben werden.
3. (2 Punkte) Führen Sie einen T-Test durch, um zu überprüfen, ob die beiden Kohorten zum Zeitpunkt T0 sich unterscheiden. Gehen Sie hierbei von normalverteilten und homoskedastischen Daten aus. Entscheiden Sie, welcher Test am sinnvollsten ist. Geben Sie in der Konsole den p-Value mit der Information aus, ob es einen signifikanten, sehr signifikanten, hoch signifikanten oder gar keinen Unterschied gibt. Nutzen Sie hierzu Ihre zuvor erstellte Signifikanz-Funktion.
4. (2 Punkte) Führen Sie einen weiteren T-Test durch, um zu überprüfen ob das Medikament einen Effekt auf die Kohorte mit erhöhtem Cholesterolspiegel zwischen beiden Zeitpunkten hat. Entscheiden Sie wieder, welcher Test hierbei der Sinnvollste ist und geben Sie in der Konsole den p-Value und die Signifikanz wieder aus.
5. (Zusatz, keine Punkte): Überprüfen Sie auch noch, ob das Medikament einen Effekt auf die Personen mit normalem Cholesterolspiegel hat und ob die beiden Kohorten zum Zeitpunkt T1 sich unterscheiden. Haben wir das Wundermittel gegen einen hohen Cholesterolspiegel gefunden?

Achtung: wenn Sie Spalten aus der Tabelle in Variablen auslesen, achten Sie darauf, dass die Daten keine NaN Werte enthalten. Sie können etwaige NaN mit folgendem Befehl löschen: `spalte = spalte.dropna()`

Aufgabe 13.2 (3 Punkte)

Drei verschiedene Kopfschmerztabletten wurden auf ihre Wirkdauer (in Minuten) an verschiedenen Personen geprüft. Nutzen Sie dazu die Daten aus der angehängten Datei „Testdaten-Mittwoch.xlsx“ aus dem Arbeitsblatt „Wirkdauer“. Sie werden feststellen, dass das die selben Daten sind wie in der zweiten Aufgabe der letzten Woche. Überprüfen Sie mittels eines ANOVA Tests, ob die Messwerte der drei Kopfschmerztabletten sich unterscheiden. Geben Sie in der Konsole den p-Value mit der Information aus, ob es einen signifikanten, sehr signifikanten, hoch signifikanten oder gar keinen Unterschied gibt. Achtung: auch hier wieder sicherstellen, etwaigen NaN Werte zu löschen.