

Praktikum 4: Testtheorie

Theoretische Aufgaben

Aufgabe 1. Jemand sagt: „Die Schülerinnen und Schüler meiner Schule sind mit 99%-iger Sicherheit 10 Prozentpunkte intelligenter als der Durchschnitt!“ Zur Begründung wird eine IQ-Stichprobe erhoben von 100 Schüler:innen, die einen mittleren IQ-Wert von 105 hat und eine Standardabweichung von 20. Geben Sie Test, Teststatistik, kritischen Wert und p -Wert an. Ist die Aussage haltbar?

Aufgabe 2. Testen Sie mithilfe des Chi-Quadrat-Tests und des G -Tests zum Signifikanzniveau von 1%, ob ein ikosaedrischer Spielwürfel gezinkt ist, bei dem folgende Häufigkeiten gewürfelt wurden.

Augenzahl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Häufigkeiten	35	52	57	41	66	70	55	34	49	49	44	56	52	61	49	51	34	50	51	44

Aufgabe 3. Entscheiden Sie, ob eine Stichprobe mit Mittelwert 0.99, Standardabweichung 1.49 und Häufigkeiten 37, 176, 676, 1644, 2516, 2442, 1625, 884 zu den Intervallgrenzen $-3, -2, -1, 0, 1, 2, 3$ normalverteilt ist.

Aufgabe 4. Entscheiden Sie, ob in Beispiel 4.6.1 der Vorlesung Unabhängigkeit vorliegt.

Programmier- und Verständnis-Aufgaben

• Empirische Verteilung

- Definieren Sie die Funktion `ecdf(x, x)`, die für eine Stichprobe x den Wert der empirischen Verteilungsfunktion F_X an der Stelle x berechnet.
- Plotten Sie nun für eine zufällig generierte standardnormalverteilte Stichprobe X die empirische Verteilungsfunktion und die Verteilungsfunktion Φ auf dem Intervall $[-5, 5]$.
- Plotten Sie außerdem den Abstand $x \mapsto F_X - \Phi_X$ beider Funktionen.
- Wie stellt sich die Abhängigkeit von der Größe der Stichprobe dar?

• Chi-Quadrat-Test

- Hier sollen ein unverfälschter und ein gezinkter Würfel, bei dem eine Augenzahl eine um 50% erhöhte Häufigkeit hat, verglichen werden.
- Dazu soll zu jeder Stichprobengröße $n = 10, 20, 50, 100, 200, 500, 1000$ und einer entsprechenden Stichprobe für den unverfälschten Würfel die Teststatistik Q_n des Chi-Quadrat-Tests für die Nullhypothese „Würfel OK“ ermittelt werden und damit Annahme oder Abweisung für das Signifikanzniveau $\alpha = 5\%$ bestimmt werden.
- Diese Stichprobentests sollen $N = 10\,000$ mal wiederholt werden, und Histogramme für die Teststatistiken (für alle n) erzeugt werden, sowie die Annahmerate berechnet werden. Plotten Sie die Histogramme und dazu zum Vergleich jeweils die Dichte der passenden Chi-Quadrat-Verteilung. Plotten Sie auch die empirische Verteilungsfunktion der Teststatistik und zum Vergleich die Verteilungsfunktion der passenden Chi-Quadrat-Verteilung. Sie sollten nun 7 mal 2 Plots erzeugt haben.
(Hinweise: Hier ist es sinnvoll den klassischen Test mit kritischer Grenze zu benutzen, denn die p -Wert-Berechnung ist langsam. Man muss nicht jeden Würfel einzeln ziehen, sondern man kann mit `rng.multinomial(n, P)` Häufigkeiten „ziehen“.)
- Führen Sie schließlich das Gleiche für Stichproben des gefälschten Würfels durch, aber unter der unveränderten Nullhypothese „Würfel OK“.
- Wie verhalten sich jeweils die Dichte-Plots für steigendes n ?
- Wie die Plots der Verteilungsfunktionen? Wie zeigt sich der Satz von Pearson?
- Wie spiegelt sich α in der Annahmerate wieder?
- Was erhalten Sie für Trennschärfen im gezinkten Fall?

• Trennschärfe von Normalitätstest

Laden Sie sich die Datei `P4vorgabe.zip` herunter und benutzen Sie das darin enthaltene Programm `main4.py`, als Grundlage für die folgenden Programmieraufgaben.

- Nutzen Sie die Bibliotheksfunktionen `kstest`, `shapiro` und `anderson` aus `scipy.stats` sowie `lilliefors` aus `statsmodels.stats.diagnostic`, um damit Funktionen

<code>KS(X, mu, sigma)</code>	(Kolmogorow–Smirnow),
<code>LF(X)</code>	(Lilliefors),
<code>AD(X)</code>	(Anderson–Darling),
<code>SW(X)</code>	(Shapiro–Wilk),

 zu definieren, die `True` oder `False` zurückgeben für Annahme bzw. Ablehnung der Nullhypothese, dass eine Normalverteilung vorliegt bei Signifikanzniveau 5%.
- Warum muss bei `KS` zusätzlich Erwartungswert und Standardabweichung übergeben werden?
- Definieren Sie nun die „Generatoren“ von Verteilungen aus `main4.py`, wie dies für `Normal(mu, sigma)` und `Uniform(a, b)` vorgemacht ist. Rückgabewert soll ein Dreitupel sein bestehend aus Stichprobengenerator, Erwartungswert und Standardabweichung für die jeweilige Verteilung.
- Nun kann `test_tests()` ausgeführt werden.
- Was wird ausgegeben, und wie ist das zu interpretieren?
- Was lässt sich über Signifikanzniveau und Trennschärfe sagen?

Abgabe

- Laden Sie das Archiv `P4vorgabe.zip` von [moodle](#) herunter, entpacken Sie es, und testen Sie Ihre Programme, indem Sie `test4.py` im gleichen Verzeichnis mit `python` ausführen. Erhalten Sie `ERROR`, so entspricht Ihr Programm nicht der Spezifikation von oben. Erhalten Sie `FAIL`, so ist Ihr Programm zwar lauffähig, aber die berechneten Werte sind fehlerhaft.
- **Abgaben, bei denen der Test gar nicht durchläuft oder mit `ERROR`, werden nicht akzeptiert**, `FAIL` führt nur zu Punktabzug.
- Sie finden in obigem Archiv auch die Datei `info4.md` mit anzugebenden Informationem zu Ihrem Team und Ihrer Abgabe, **bitte füllen Sie diese nach dortiger Anleitung aus, und vergessen Sie nicht die Quellenangabe**. Abgaben mit unvollständiger Datei `info4.md` können nicht gewertet werden. **Belassen Sie diese Datei in der UTF-8-Kodierung**.
- Komprimieren und bündeln Sie alle oben erzeugten oder geänderten Dateien, indem Sie ein ZIP-Archiv erstellen. Sollten Sie nicht wissen, wie das geht, konsultieren Sie dazu die Dokumentation Ihres Betriebssystems.
- Nennen Sie Ihr ZIP-Archiv `P4.zip`.
- Schreiben Sie eine Email an rosehr@hm.edu mit Betreff **Abgabe Datenanalyse**, Dateianhang **P4.zip** und irgendwelchem sonstigen Inhalt. **Der Automat akzeptiert die Abgabe nur, wenn diese Angaben (Betreff, Dateiname), sowie Ihre persönlichen Angaben (Nach-, Vor- und Teamname) korrekt sind**. Sie erhalten innerhalb der nächsten 24 Stunden (meistens aber deutlich schneller) eine Bestätigungsemail, wenn alles korrekt war.