

1. Aufgabe Sie wollen zwei Hirsesorten hinsichtlich ihrer Trockenempfindlichkeit vergleichen. Dazu werden die Pflanzen 7 Tage lang nicht gegossen und dann ihre mittlere Blattfläche bestimmt. Wie viele Pflanzen von jeder Sorte brauchen Sie für den Versuch, wenn ein Blattflächenunterschied von 20 auf dem Signifikanzniveau 5% bei einer Testmacht von 80% erkannt werden soll? Verwenden Sie dazu, dass die Standardabweichungen in einem Vorversuch 29.55 und 60.21 waren.

a) Berechnen Sie die benötigte Stichprobenlänge in jeder Gruppe ohne Verwendung von R. Verwenden Sie dazu die Verteilungstabellen von der Vorlesungs-Homepage.

b) Berechnen Sie die benötigte Stichprobenlänge in jeder Gruppe mit R.

c) Angenommen Sie hätten die benötigte Stichprobenlänge nicht bestimmt und sich aus dem Bauch heraus für 20 Pflanzen pro Sorte entschieden. Wie groß ist dann die Testmacht? Drücken Sie diese Testmacht in Worten als relative Häufigkeit aus.

2. Aufgabe Ein medizinischer Test auf eine neue Grippe schlägt bei 97 % aller Infizierten an und fälschlicherweise auch bei 2% der Nichtinfizierten. Obwohl ich weiß, dass unter tausend Erkälteten nur einer mit der neuen Grippe infiziert ist, mache ich mir wegen meiner laufenden Nase Sorgen und lasse mich testen. Falls der Test anschlägt, wie groß ist dann die Wahrscheinlichkeit, dass ich mit der neuen Grippe infiziert bin?

3. Aufgabe In den Flüssen Trinidads sind die Guppys *Poecilia reticulata* verschiedenen Räubern ausgesetzt. Flussabwärts gibt es den großen und gefährlichen *Crenicichla alta*, flussaufwärts den weniger gefährlichen *Rivulus hartii*. Evolutionär-ökologische Überlegungen führen zu der Hypothese, dass die Nachkommen flussabwärts kleiner sind. Um diese Hypothese zu prüfen, werden Sie nach Trinidad fliegen, um das Gewicht von neugeborenen Guppys zu messen. Wie viele neugeborene Guppys müssen Sie vermessen, wenn ein tatsächlicher mittlerer Gewichtsunterschied von 0.4 mg auf dem Signifikanzniveau 5% erkannt werden soll? Die Testmacht soll 99% Prozent sein, da die Messreihe wegen den Reisekosten recht teuer ist. Aus Voruntersuchungen wissen Sie, dass Sie in den Daten mit einer Standardabweichung von 0.5 mg rechnen können.

a) Ermitteln Sie die benötigte Stichprobenlänge an den beiden Flussabschnitten ohne Verwendung von R. Verwenden Sie dazu die Verteilungstabellen von der Vorlesungs-Homepage.

b) Berechnen Sie die benötigte Stichprobenlänge an den beiden Flussabschnitten mit R.

c) Angenommen, die Messreihe würde – rein hypothetisch – 100 mal durchgeführt werden. In wie vielen Fällen erwarten Sie bei einer Testmacht von 99%, dass die Daten keine Signifikanz liefern, wenn der wahre mittlere Gewichtsunterschied tatsächlich 0.4 mg beträgt.

d) Angenommen Sie hätten die benötigte Stichprobenlänge nicht bestimmt und sich aus dem Bauch heraus für 20 Neugeborene pro Flussabschnitt entschieden. Wie groß ist dann die Testmacht? Drücken Sie diese Testmacht in Worten als relative Häufigkeit aus.

4. Aufgabe Experimente müssen sorgfältig durchdacht werden, um Fehler zu vermeiden.

a) Im Raum München wurden im Laufe eines Tages 30 Eichhörnchen (*Sciurus vulgaris*) gefangen und durchnummeriert zum Zwecke eines Verhaltensexperimentes. Wie sollen die 30 Tiere auf zwei Beobachtungsgruppen aufgeteilt werden?

b) Sicherlich hat die jahreszeitliche Temperatur einen Einfluss auf das Nuss-Sammelverhalten der Eichhörnchen. Nun möchte man herausfinden, ob auch die Zusammensetzung des Lichtes einen Einfluss auf das Sammelverhalten hat. Die Idee ist, die eine Gruppe blauem Licht ausgesetzt und die andere Gruppe rotem Licht. Wie könnte der Versuch konkret ablaufen? Welche Fehler könnten gemacht werden, welche Einflussgrößen könnten das Ergebnis verfälschen?

5. Aufgabe (*t*-Statistik und Permutationstests) Anstatt zur Bestimmung des *p*-Werts (approximative) Aussagen über die Verteilung der *t*-Statistik zu benutzen, kann man die Hypothese „die beiden Stichproben stammen aus derselben Population“ auch mittels eines Permutationstests prüfen, beispielsweise folgendermaßen: Verteile die Gruppenbezeichnungen rein zufällig (Hinweis: der R-Befehl `sample(x)` erzeugt eine zufällige Permutation des Vektors x), berechne die *t*-Statistik für die so permutierten Daten (mit `t.test()$statistic` und bestimme durch wiederholte Simulation, mit welcher Wahrscheinlichkeit sich dabei ein (betragsmäßig) mindestens so großer Wert der *t*-Statistik wie der tatsächlich beobachtete ergibt. Dieses Vorgehen ist zwar aufwändiger als der klassische *t*-Test, dafür aber „immun“ gegen Verletzungen der Normalverteilungsannahme. Führen Sie den Permutationstest wie gerade beschrieben für die Milben-Daten aus `milben.csv` durch, gemäß dem Beispiel aus der Vorlesung über den zwei-Stichproben-*t*-test. Vergleichen Sie dann das Ergebnis mit dem *p*-Wert des *t*-Tests.

6. Aufgabe Um den Anteil $p \in [0, 1]$ von Personen in einer (großen) Gruppe, die eine „heikle“ Eigenschaft haben, zu schätzen, betrachten wir folgendes Verfahren: Als Teilnehmer erhalten Sie die Frage, beispielsweise „Haben Sie schon einmal Marihuana geraucht?“ und zwei Würfel (mit je sechs Seiten). Sie würfeln verdeckt: Wenn die Augensumme 5 oder 6 ist, antworten Sie „ja“, wenn sie 8 oder 9 ist, antworten Sie „nein“, ansonsten beantworten Sie die Frage.

a) Wenn der tatsächliche Anteil der Personen, die die heikle Eigenschaft haben, p ist, wie wahrscheinlich ist es dann, dass eine rein zufällig herausgezogene Person mit „ja“ antwortet?

b) Wie könnte man anhand der Ergebnisse der Befragung von n Personen p schätzen? Bestimmen Sie auch die Varianz für Ihren Schätzer (durch Rechnung oder durch Simulation für verschiedene Wahlen von n und p).