

ÜBUNGEN ZUR WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK FÜR BIOLOGEN

Blatt 11

Es ist empfehlenswert, ein Übersicht über alle statistischen Tests aus der Vorlesung anzufertigen. Diese Übersicht wird bei einigen Aufgaben und später in der Klausur hilfreich sein.

1. Aufgabe Kuss et al. ("The fouled player should not take the penalty himself": An empirical investigation of an old German football myth, *J. Sports Sciences* **25**, no. 9, 963–967, 2007) berichten über die Strafbüchse in der 1. Fußballbundesliga (der Herren) von August 1993 bis Februar 1995:

	verwandelt	nicht verwandelt
der Gefoulte schießt selbst	74	28
anderer Spieler schießt	547	186

Stützen diese Daten die These, dass der Gefoulte den Elfmeter nicht selbst schießen sollte?

- Formulieren Sie zunächst die Nullhypothese.
- Führen Sie dann einen geeigneten Test „per Hand“ durch, d.h. ohne Verwendung eines R-Befehl, in dem „test“ vorkommt. Verwenden Sie dazu die Verteilungstabellen von der Vorlesungs-Homepage.
- Führen Sie dann einen geeigneten Test mit R durch.

2. Aufgabe Es wird vermutet, dass bei einem Pferderennen auf einem ovalen Parcours die Startposition einen Einfluß auf die Gewinnchance hat. In 144 Rennen hatten die Sieger die Startpositionen 1, 2, . . . , 8 mit den folgenden Häufigkeiten: 29, 19, 18, 25, 17, 10, 15, 11. Testen Sie die Hypothese, dass die Startposition keinen Einfluß auf die Siegchance hat, zum Irrtumsniveau 5%.

3. 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 15 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.

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

4. 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.

5. 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?

6. Aufgabe (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 von Blatt 6, Aufgabe 3 durch. Vergleichen Sie dann das Ergebnis mit dem p -Wert des t -Tests.