ÜBUNGEN ZUR WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK IN DER BIOLOGIE

Blatt 12

Aufgabe 1 Es könnte sein, dass mit Chlor versetztes Wasser (wie in Schwimmbädern üblich) den Zahnschmelz beeinträchtigt. Dazu wurden 200 Schwimmer je nach dem, ob sie mehr oder weniger als 6 Stunden wöchentlich trainieren, in zwei Gruppen eingeteilt und der Zahnschmelz untersucht, mit folgendem Ergebnis:

	Zahnschmelz angegriffen		
Schwimmzeit pro Woche	Ja	Nein	Gesamt
mehr als 6 h	29	71	100
weniger als 6 h	19	81	100
Gesamt	48	152	200

a) Stützen die Beobachtungen die Hypothese des ersten Satzes der Aufgabenstellung? Formulieren Sie eine Nullhypothese, testen Sie mittels eines χ^2-Tests und formulieren Sie einen Antwortsatz.

b) Schätzen Sie den Anteil θ der Personen mit angegriffenem Zahnschmelz unter den Viel-Schwimmern. Geben Sie auch ein 95%-Konfidenzintervall an.

c) Nehmen wir an, analoge Ergebnisse wären in einer Studie des doppelten Umgangs beobachtet worden: Je 200 Viel- und 200 Wenig-Schwimmer wurden untersucht, 58 der Viel- und 38 der Wenig-Schwimmer hatten angegriffenen Zahnschmelz. Welchen Wert würde die χ^2-Statistik dann annehmen, und wie lautete dann Ihr Befund?

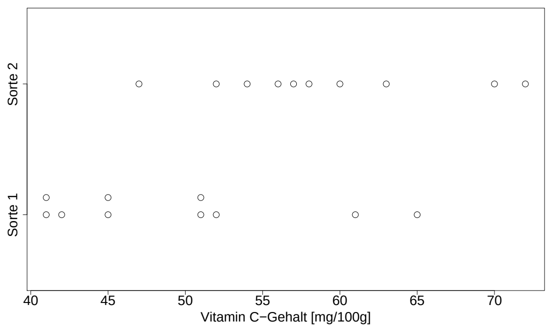
d) Welches Konfidenzintervall für θ würde sich mit den Daten aus c) ergeben?

Aufgabe 2 In einer Studie wurde die Wirksamkeit eines Schlafmittels geprüft und für zehn Patienten die zusätzliche Anzahl Stunden Schlaf pro Tag (gemittelt über den Zeitraum der Medikamentengabe) im Vergleich zu einem Referenzzeitraum ohne Medikament bestimmt:

1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4

Der Stichproben-Mittelwert der zusätzlichen Stunden Schlaf ist $\mu = 2.33$ mit Stichproben-Varianz $s^2 = 4.01$. Testen Sie (unter einer Normalverteilungsannahme an die Daten) die Wirksamkeit des Schlafmittels. Hinweise: Wie immer zuerst die Nullhypothese formulieren. Geben Sie auch ein 95%-Konfidenzintervall für die mittlere zusätzliche Anzahl Stunden Schlaf mit Medikament an.

Aufgabe 3 Der Vitamin C-Gehalt zweier Kohlsorten wurde in jeweils 10 Stichproben pro Sorte untersucht:



Dabei wurde für die Proben von Sorte 1 ein mittlerer Gehalt $\mu_1 = 49.4$ mg pro 100g mit Standardabweichung $s_1 = 8.33$ gefunden, für Sorte 2 waren die Werte $\mu_2 = 58.9$ und $s_2 = 7.74$. Testen Sie die Hypothese, dass der mittlere Vitamin C-Gehalt für beide Sorten gleich ist, mittels eines t-Tests zum Irrtumsniveau 5% und formulieren Sie Ihr Ergebnis in einem Satz. Nehmen Sie dazu an, dass die wahren Varianzen gleich sind.

Aufgabe 4 Phenylketonurie ist eine autosomal-rezessiv vererbte menschliche Stoffwechselkrankheit (d.h. nur homozygote Träger des „kranken“ Allels a erkranken), an der (in Deutschland) ca. eines von 8000 Neugeborenen leidet.
Nehmen wir an, die Population befinde sich (bezüglich dieses Gens) im Hardy-Weinberg-Gleichgewicht. Welchen Anteil der Population haben dann die Genotypen AA, aA und aa?
Ein gesundes Paar (gebildet gemäß der „Hardy-Weinberg-Regeln“) habe ein gesundes Kind. Wie wahrscheinlich ist es, gegeben diese Information, dass beide Eltern Genotyp AA haben?

Aufgabe 5 (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 (Hinweis: `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.

Aufgabe 6 Für ein Gen, für das vier Allele A, B, C und D unterscheidbar sind, wurden in einer zufälligen Stichprobe von 100 Individuen die verschiedenen Genotypen in folgenden Anzahlen gefunden:

AA	AB	AC	AD	BB	BC	BD	CC	CD	DD
26	11	12	9	10	13	7	4	5	3

(a) Berechnen Sie das 95%-Wald-Konfidenzintervall für die Häufigkeit des Allels A in der Population.

(b) Es stellt sich die Frage, ob sich die Population im Hardy-Weinberg-Gleichgewicht befindet. Führen Sie hierzu einen geeigneten χ^2-Test durch (inklusive Nullhypothese und Antwortsatz). Sie dürfen dabei verwenden, dass der Wert der Teststatistik gleich $X^2 = 14.32$ ist. Erläutern Sie aber, was der Genotyp AA zu diesem X^2-Wert beiträgt.