

Mathematik II für Biologen

Übungsblatt 12 (Abgabe am 10.07.2015)

Aufgabe 44

(10 Punkte)

Eine räumlich homogene Ackerfläche wurde in 100 gleich große Quadrate eingeteilt. In der Mitte eines jeden Quadrates wurde eine Falle aufgestellt. Wir bezeichnen mit x_i die Anzahl Exemplare des Laufkäfers *CARABUS GRANULATUS*, die nach einer gewissen Zeit in der i -ten Falle gefunden wurden ($i = 1, 2, \dots, 100$). Die folgende Tabelle gibt die Anzahl $n(x)$ der Fallen mit genau x solchen Käfern an.¹⁶

x	0	1	2	3	4	5	6	7	8	9	10	≥ 11	Total
$n(x)$	22	17	24	10	14	2	1	3	3	3	1	0	100

Wir nehmen als (Null-)Hypothese H_0 an, dass die Käfer „gleichmäßig“ über das Feld verteilt sind und sich unabhängig voneinander bewegen.

- a) Warum ist die Anzahl von Käfern in jeder Falle in etwa poissonverteilt, wenn H_0 stimmt?

Laut Vorlesung sind der Erwartungswert $E[X]$ und die (theoretische) Varianz $\text{Var}(X)$ einer poissonverteilten Zufallsvariable mit Parameter λ beide gleich λ . Wenn obige Daten $x_1, \dots, x_{100}$ also von unabhängigen poissonverteilten Zufallsvariablen stammen, so sollten auch ihre empirischen Entsprechungen, nämlich der arithmetische Mittelwert $\bar{x}$ und die empirische Varianz s_x^2 , in etwa übereinstimmen, d.h. der Quotient $s_x^2/\bar{x}$ sollte in etwa gleich 1 sein.

- b) Berechnen Sie (ähnlich wie in Aufgabe 2) den empirischen Mittelwert $\bar{x}$ und die empirische Varianz s_x^2 der Stichprobe $x_1, \dots, x_{100}$ sowie den Quotienten $s_x^2/\bar{x}$, entweder per Hand und mit dem Taschenrechner (Rechenweg angeben) oder mit MATLAB (Code angeben und erläutern).

Nun kann man natürlich auch bei perfekter Übereinstimmung der Daten mit dem Modell nicht erwarten, dass der empirische Quotient $s_x^2/\bar{x}$ genau gleich 1 ist. Er wird nach wie vor zufällig um 1 herum schwanken. Nur wie stark?

- c) Um dies zu untersuchen, erzeugen Sie bitte 100 poissonverteilte Zufallszahlen mit Parameter λ , wobei Sie als λ den Wert $\bar{x}$ nehmen, den Sie in Aufgabe b) berechnet haben, und berechnen Sie zu dieser simulierten Stichprobe den empirischen Quotienten $s_x^2/\bar{x}$ (=Teststatistik). Wiederholen Sie das Ganze 1000 Mal und tragen Sie die so erhaltenen Quotienten in ein Histogramm ein. MATLAB-Code:

```
pois=poissrnd(???, 100, 1000) % ??? = arithm. Mittelwert der echten Daten
quotienten=var(pois)./mean(pois);
hist(quotienten)
```

Wo befindet sich der Quotient $s_x^2/\bar{x}$ der echten Daten in diesem Histogramm? Wie lautet demnach in etwa der p-Wert für diesen Test? Was schließen Sie daraus für H_0 ?

¹⁶Quelle: G. Welzl, F. P. Wahl. Statistische Modelle zur Erfassung der räumlichen Verteilung von Laufkäfern in einer Agrarlandschaft. Tagungsberichte der Arbeitsgruppe Ökologie, Dt. Region der Internat. Biometr. Ges., Heft 6, Mai 1995.

Aufgabe 45

(10 Punkte)

Die folgende Tabelle enthält die Gewichte von 10 Braunviehkühen in kg direkt vor dem Alpauftrieb und in der vierten Alpwoche. (Quelle: R. E. Messikommer. Schätzung des Futtermittelfressens von Kühen auf der Alpweide. Diss. ETH 8282)

	Mona	Britta	Roswita	Nora	Palme	Sorina	Elbe	Gallina	Carla	Grille
vor der Alpung	634	576	603	516	535	550	556	646	560	664
in der 4. Alpwoche	565	545	541	489	514	538	516	595	528	635
Gewichtsabnahme										

- Bestimmen Sie ein beidseitiges 95%-Vertrauensintervall $[a; b]$ im Sinne eines z -Tests für die erwartete Gewichtsabnahme. Nehmen Sie dazu an, dass die Gewichtsabnahmen der einzelnen Kühe unabhängig voneinander und identisch normalverteilt sind und dass die tatsächliche (theoretische) Standardabweichung σ gleich der empirischen Standardabweichung ist.
- Leiten Sie daraus ein 95%-Vertrauensintervall her für die Wahrscheinlichkeit p , dass die Kuh Karoline, die auch auf der Alp ist, aber nicht gewogen wurde, zugenommen hat. (Anleitung: Berechnen Sie dazu zu den beiden Extremwerten a und b jeweils das zugehörige p . Diese beiden Werte bilden die Grenzen des Vertrauensintervalls für p .)
- Bestimmen Sie das 95%-Vertrauensintervall für die in (b) genannte Wahrscheinlichkeit p im Sinne eines zweiseitigen Binomialtests. (Bitte keine Faustregel anwenden, da $n = 10$ zu klein ist.)
- Warum ist das in c berechnete Intervall viel größer als das in b berechnete?
HINWEIS: Welche Annahmen wurden gemacht?

Aufgabe 46

(10 Punkte)

Die bereits aus den Aufgaben 38 und 43 bekannte Datei `fishy.dat` enthält in der ersten Spalte die Länge (in cm), in der zweiten Spalte das Gewicht (Einheit leider unbekannt) und in der dritten Spalte den DDT-Gehalt (in ppm) von $n = 96$ Welsen, die im Tennessee River in Alabama, USA, gefangen wurden. Bearbeiten Sie das Problem aus Aufgabe 43 a mit Hilfe eines zweiseitigen

- t-Tests,
- Wilcoxon-Tests,

d.h. testen Sie mit diesen Tests, ob die Daten mit der Annahme verträglich sind, dass das durchschnittliche Gewicht μ eines Welses aus diesem Fluss gleich 920 ist. MATLAB-Code:

```
>> load fishy.dat
>> gewicht=fishy(:,2)
>> [h,p,vi]=ttest(gewicht,920)
>> [p,h]=signrank(gewicht,920) % = Wilcoxon-Test
```

Interpretieren Sie die Ergebnisse.

Aufgabe 47

(10 Punkte)

Um den Umgang von jungen und älteren Urlaubern am Strand mit der Gefahr von Sonnenbrand zu analysieren, wurde eine Gruppe von insgesamt N Personen nach einem Tag am Meer untersucht. Untenstehende Tabelle zeigt, welcher Anteil der N Personen den Strand mit oder ohne Sonnenbrand verlassen hat und in die untere oder in die obere Altersklasse gehört.

Alter \ Hautzustand	Sonnenbrand	KEIN Sonnenbrand
15-25 Jahre	30%	20%
über 25 Jahre	20%	30%

- Nehmen Sie $N = 50$ an und testen Sie die Nullhypothese H_0 : *Alter und Hautzustand sind unabhängig* mit einem χ^2 -Test auf dem Niveau 5%. Kann H_0 verworfen werden?
- Wie groß hätte N mindestens sein müssen, damit H_0 mit denselben Prozentzahlen auf dem Niveau 5% verworfen worden wäre?