

Mathematik II für Biologen

Übungsblatt 11 (Abgabe am 10.07.2009)

Aufgabe 36

(20 Punkte)

→ Bitte bearbeiten Sie diese Aufgabe ohne MATLAB o.ä., Taschenrechner ist o.k. ←
Der Fettgehalt der Milch von neun verschiedenen Kühen derselben Rasse beträgt in Prozent

4,59 4,48 4,21 4,24 4,45 4,35 4,08 4,03 4,63.

- a) Testen Sie mit einem **z-Test** unter der Annahme, dass die “wahre” Standardabweichung σ des Fettgehaltes 0,2 beträgt, auf dem 5%-Niveau die Nullhypothese, dass der “wahre” mittlere Fettgehalt $\mu = 4,22\%$ ist, gegenüber den Alternativen

(i) $\mu \neq 4,22\%$

(ii) $\mu > 4,22\%$.

Geben Sie hierbei die konkrete Formel für die Teststatistik an, deren Verteilung unter Annahme der Nullhypothese, den jeweiligen Verwerfungsbereich, den beobachteten Wert der Teststatistik und das jeweilige Testergebnis. Geben Sie außerdem jeweils das zugehörige 95%-Vertrauensintervall für μ an.

- b) Testen Sie mit einem **t-Test** auf dem 5%-Niveau die Nullhypothese, dass der theoretische, “wahre” mittlere Fettgehalt $\mu = 4,22\%$ ist, gegenüber der Alternative

(i) $\mu \neq 4,22\%$

(ii) $\mu > 4,22\%$.

Geben Sie hierbei die konkrete Formel für die Teststatistik an, deren Verteilung unter Annahme der Nullhypothese, das jeweilige Verwerfungskriterium, den beobachteten Wert der Teststatistik und die jeweilige Testentscheidung. Geben Sie außerdem jeweils das zugehörige 95%-Vertrauensintervall für μ an.

- c) Testen Sie mit einem **Wilcoxon-Test** auf dem 5%-Niveau die Nullhypothese, dass der theoretische, “wahre” mittlere Fettgehalt $\mu = 4,22\%$ ist, gegenüber der Alternative

(i) $\mu \neq 4,22\%$

(ii) $\mu > 4,22\%$.

Berechnen Sie dazu die Teststatistik (Rechenweg zeigen!), bestimmen Sie das jeweilige Verwerfungskriterium und formulieren Sie die jeweilige Testentscheidung.

- d) Mit welcher Wahrscheinlichkeit kann man mit dem **z-Test** aus Aufgabenteil a (ii) und $n = 9$ Daten statistisch auf dem Signifikanz-Niveau $\alpha = 5\%$ zeigen, dass $\mu > 4,22\%$ ist, wenn in Wirklichkeit $\mu = 4,4\%$ ist? (Dies ist die Macht des Tests, $1 - \beta$, für die Nullhypothese $H_0 : \mu = 4,22$ gegenüber der Alternative $H_A : \mu = 4,4\%$.)

HINWEIS: Erinnern Sie sich daran, dass in Teil a (ii) H_0 genau dann verworfen wird, wenn

$$Z := \frac{S_n - 4,22\%n}{\sigma\sqrt{n}} > 1,64, \quad \text{wobei} \quad S_n := X_1 + \dots + X_n \quad \text{ist.}$$

Lösen Sie diese Ungleichung nach S_n auf. Wie ist, andererseits, S_n verteilt, wenn H_A gilt?

Aufgabe 37

(10 Punkte)

Die bereits aus Aufgabe 31 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.

- a) Testen Sie mit Hilfe eines zweiseitigen
 - (i) Vorzeichen-Tests,
 - (ii) z -Tests,ob die Daten mit der Annahme verträglich sind, dass das durchschnittliche Gewicht μ eines Welses aus diesem Fluss gleich 920 ist. Bestimmen Sie auch jeweils ein 95%-Vertrauensintervall für μ . Nehmen Sie hierbei für den z -Test an, dass die theoretische ("wahre") Standardabweichung σ gleich 250 ist.
- b) Damit man einen z -Test überhaupt anwenden darf¹⁰, sollten die Daten halbwegs durch eine Normalverteilung beschrieben werden können. Zeichnen Sie, um diese Annahme zu überprüfen,
 - (i) einen Q-Q-Plot (siehe Aufgabe 8) der Gewichtsdaten und
 - (ii) die (empirische kumulative) Verteilungsfunktion der Daten zusammen mit der (theoretischen kumulativen) Verteilungsfunktion der Normalverteilung, wobei Sie als Parameter der Normalverteilung (Erwartungswert und Varianz) die aus den Daten geschätzten empirischen Entsprechungen (Mittelwert und empirische Varianz) nehmen.
HINWEIS: MATLAB erwartet als drittes Argument von `normcdf` die Standardabweichung, OCTAVE dagegen die Varianz.Zeichnen Sie auch die entsprechenden Diagramme für die DDT-Werte.¹¹ Welcher der beiden Datensätze lässt sich demnach besser durch eine Normalverteilung beschreiben? Warum?
- c) Wieviel Welse müsste man insgesamt fangen und wiegen, damit das 95%-Vertrauensintervall für den z -Test der gemessenen Gewichte weniger als 50 lang ist, vorausgesetzt, σ ist weiterhin gleich 250?

MATLAB-Code (unvollständig):

```
>> load fishy.dat
>> gewicht=fishy(:,2)
>> [h,pwert,vi]=ztest(gewicht,920,250) % siehe help ztest
>> qqplot(gewicht)
>> cdfplot(gewicht)
>> a=min(gewicht);b=max(gewicht);x=a:(b-a)/100:b;
>> hold on
>> plot(x,normcdf(x,???,???) % siehe help normcdf
>> hold off
```

¹⁰Das gleiche gilt für t -Tests.

¹¹Hier lohnt es sich von Hand geeignete Achsenabschnitte zu wählen, um die Daten beurteilen zu können; MATLAB-Befehl `axis([??? ??? -Inf Inf])` (siehe MATLAB-Hilfe) oder durch Klicken.